

**SYSTEMATISCHE FEHLER
ALS FOLGE DES STICHPROBENFEHLERS
BEI VERSCHIEDENEN FAKTORENANALYTISCHEN
VERFAHREN, UND EIN NEUER WEG
ZUR BESTIMMUNG DER FAKTORENANZAHL**

**Inaugural - Dissertation
zur
Erlangung der Doktorwürde
der Philosophischen Fakultäten
der Albert-Ludwigs-Universität
zu Freiburg i.Br.**

**vorgelegt von
Ulrich Schiller
aus Magdeburg**

1979

Referent: Prof. Dr. J. Fahrenberg
Korreferent: Prof. Dr. W. Bauer
Sprecher: Prof. Dr. H. Steger
Tag der Promotion: 15. Dezember 1978



SIMPLICITY IS ELEGANCE; ELEGANCE IS SIMPLICITY.

But the most important future work, as I see it, should continue to concentrate on the number-of-factors question.

H.F.Kaiser "A second generation little Jiffy" 1970

I.	<u>Vorwort</u>	1
II.	<u>Ansatzpunkte zu dieser Untersuchung</u>	
1.	Ein erster Überblick	3
2.	Validitätskoeffizienten aus Simulations - Studien	4
3.	Fehler - Faktoren	9
4.	Multiple Korrelation und die Kommunalitätsschätzung	12
5.	Measure of Sampling Adequacy (MSA)	
5.1.	Exkurs zur Image - Theorie	16
5.2.	Die Beurteilung der Stichprobenadäquatheit	18
6.	Zusammenfassung und erste Zielsetzung	21
III.	<u>Betrachtungen zum Grundkonzept der Faktorenanalyse</u>	
1.	Die Aufspaltung der Einheitsvarianz	24
2.	Die Bedingungen der Null - Kovarianzen	25
3.	Schlußfolgerungen	28
4.	Folgerungen für die Auswertung	
4.1.	Gesamtstichprobenebene	29
4.2.	Ebene der einzelnen Variablen	30
IV.	<u>Die Ergebnisse</u>	
1.	<u>Der systematische Fehler auf der Stichprobenebene</u>	
1.1.	Die Bedeutung der Versuchspersonenanzahl	
1.1.1.	Überblick über die Ergebnisse bei den einzelnen Analysenverfahren	33
1.2.	Die Bedeutung der mittleren gemeinsamen Varianz	42
1.2.1.	Varianzausschöpfung in den einzelnen Analysen- verfahren	42
1.2.2.	Der systematische Fehler in Abhängigkeit von der mittleren Input - Kommunalität (TFA1 und IFA)	45
1.2.3.	Der systematische Fehler in Abhängigkeit von der mittleren Input - Kommunalität (IFA und SMC)	47
1.2.4.	Der systematische Fehler in Abhängigkeit von der mittleren Input - Kommunalität (TFAK, AFA und KFA)	50
1.2.5.	Der systematische Fehler in den verschiedenen Analysenverfahren bei 15 Variablen	51

1.3.	Der systematische Fehler als Funktion von gemeinsamer Varianz und Versuchspersonenanzahl	52
1.3.1.	Vorbemerkungen	
1.3.2.	Bias der Hauptkomponentenanalyse (TFA1)	53
1.3.3.	Bias der Image - Analyse (IFA)	54
1.3.4.	Bias der Kanonischen Analyse (KFA)	55
1.3.5.	Bias der Quadrierten Multiplen Korrelationen (SMC)	56
1.3.6.	Die Bias - Darstellungen 15 bis 18	57
1.4.	Der systematische Fehler der $\bar{SMC}$	
1.4.1.	Die Zielsetzung	58
1.4.2.	Die allgemeine Formel	58
1.4.3.	Die allgemeine Formel für beliebige Variablenanzahl	60
1.4.4.	Diskussion der allgemeinen Gleichung	64
1.4.5.	Die Grenze für sichere Lösungen	67
2.	<u>Der systematische Fehler auf der Einzelvariablenebene - Der Nivellierungsfehler -</u>	
2.1.	Vorbemerkungen	70.
2.2.	Beispiel einer Stichprobe	71
2.3.	Die Übereinstimmung von Input- und Output-kommunalitäten	73
2.4.	Der Nivellierungsfehler	77
2.5.	Konsequenzen aus dem Nivellierungsfehler	85
2.6.	Eine Ableitung des Nivellierungsfehlers	85
V.	<u>Ein neues Verfahren zur Bestimmung der Faktorenanzahl</u>	
1.	Vorbemerkungen	89
2.	Das Verfahren	90
VI.	<u>Anwendungen</u>	
1.	Anwendungsbeispiele aus den simulierten Daten	103
2.	Ein Anwendungsbeispiel aus der Literatur Lienerts Regressionshypothese	108
3.	Ein Anwendungsbeispiel aus der Literatur Faktorenstruktur des FPI	113
VII.	<u>Zusammenfassung</u>	117
VIII.	<u>Anhang</u>	
1.	Die Daten	118
2.	Auswertungsbeispiele	122
3.	Beispiele von Populationsfaktorladungsmatrizen	130
IX.	<u>Literatur</u>	131

I. VORWORT

Diese Arbeit basiert auf den umfangreichen Vorarbeiten, vor allem im Hinblick auf die Erstellung eines Computer-Programmes, die Herr Wittmann und ich in den Jahren 1968 bis 1972 für die Diplomarbeit gemeinsam geleistet haben. Die Anregung zu dieser Arbeit kam von Herrn Professor Dr. W. Bauer. Damals war das Ziel, die Hauptkomponentenanalyse mit der Image-Analyse zu vergleichen. Grundlage des Vergleichs waren Validitätskoeffizienten aus der Korrelation ursprünglich generierter Faktorwerte mit den nach der Analyse geschätzten Faktorwerten. Dieser Versuchsplan entspricht in seinen wesentlichen Zügen dem von Oberla (1968). Oberlas Transformationsmatrix wurde allerdings von uns (Schiller, Wittmann 1971) erweitert, um Simulationen mehrdimensionaler Faktorenräume und die Verletzung der "simple structure"-Kriterien zu ermöglichen.

Eine Überlegenheit eines der untersuchten Analysenverfahren wurde letzten Endes nicht deutlich erkennbar. Dieses Ergebnis gilt besonders für kritische Bereiche, zum Beispiel für die Bestimmung der Faktorenanzahl bei niedrigen Populationsfaktorladungen oder bei geringer Versuchspersonenanzahl.

Der erhebliche Zeitaufwand für die damalige Diplomarbeit und das umfangreiche Datenmaterial brachten uns zu der Entscheidung, hier eine Basis für jeweilige Dissertationen zu sehen. Dabei sollte je ein zusätzliches Verfahren, Alpha-Analyse (Kaiser u.Caffrey, 1966) bzw. Kanonische Analyse (Rao, 1955) in die Untersuchung mit einbezogen werden.

Inzwischen hat mir Herr Wittmann dankenswerter Weise einen Teil seiner "Alpha"-Ergebnisse mit zusätzlichen Simulationen zur Verfügung gestellt. Damit können in der vorliegenden Arbeit

TFA1 = Hauptkomponentenanalyse mit 1 in der Diagonalen der analysierten Korrelationsmatrix

TFAK = Hauptachsenanalyse mit SMC-Werten als Kommunalitätenschätzung

IFA = Image - Analyse (Guttman 1953)

AFA = Alpha Faktorenanalyse

KFA = Kanonische Faktorenanalyse

mit einbezogen werden.

Die ersten Schritte und Versuche, das vorliegende Datenmaterial mit anderen Techniken als allein der Validitätskoeffizientenbestimmung zu bearbeiten, ergaben zwei Leitsätze für die Zielsetzung bei der Auswertung der Daten für die vorliegende Arbeit:

- A. Möglichst hohe direkte Auswertbarkeit und Anwendbarkeit der Ergebnisse für den Praktiker
- B. Die Vorkenntnisse über die Faktorenstruktur in der Population und in der Stichprobe sollen es ermöglichen, rückwirkend aus den Analysenergebnissen Aussagen zur "wahren Struktur" zu gewinnen.

Im Rückblick kann gesagt werden, daß diese Ziele erreicht wurden. Das gilt besonders für die während der umfangreichen Auswertung der Daten gehegte Vorstellung, Varianz-Überschätzungen bzw. Unterschätzungen in den jeweiligen Analysenverfahren in Abhängigkeit von den Kennwerten der Stichprobe zu bestimmen.

Mein besonderer Dank gilt Herrn Professor Dr. Fahrenberg, der die Betreuung dieser Arbeit übernommen hat. .

II. ANSATZPUNKTE ZU DIESER UNTERSUCHUNG

1. Ein erster Überblick

Der hier in Form einer Ebenenaufteilung erstellte Überblick ist gleichzeitig ein Ablaufplan. Er entspricht dem Oberlaschen Ablaufschema der Modellrechnung (Oberla 1968, S.274). Eine Änderung liegt nur durch die Verwendung verschiedener faktorenanalytischer Verfahren vor.

Unabhängiger Bereich	Abhängiger Bereich	Daten	Ebene
Faktorwerte Populationsfaktorlad. Anzahl der Vpn (N) Anzahl der Var (n) Anzahl der Fakt. (k)		1	1.Ebene: (<u>Population</u>)
Simulation	Stichprobenfaktorl. Korrelationsmatrix	1	2.Ebene: <u>Stichprobe</u>
	Multiple Korrelation SMC-Kommunalitäten	1	Zwischen- Ebene
Kommunalitäten- schätzung (2 Arten) Analysenverfahren (4 Formen)	Faktorladungs- matrix nach der Analyse	5	3.Ebene: <u>Analyse</u>
Rotationsverfahren (1 Verfahren)	Faktorladungs- matrix nach der Rotation	5	4.Ebene: <u>Rotation</u>
Faktorwerte-Schätzung			5.Ebene: Validitäts- bestimmung

Oberla bezieht sich bei seiner Überprüfung der Faktorenanalyse über die Validitätsberechnung auf die 1. und 4. Ebene. Der Ansatz hier geht nun davon aus, daß in den restlichen Ebenen noch viele

Informationen auszuschöpfen sind. Dies erforderte eine Entscheidung im Hinblick auf die Vergleichsmöglichkeiten der verschiedenen Ebenen. Die jeweiligen Kommunalitäten lassen solche Ebenen-Vergleiche zu. Mit dieser Entscheidung für die schwerpunktmäßige Betrachtung der Kommunalitäten ist die 4. Ebene, das Rotationsproblem, nicht mehr einbeziehbar. Die Kommunalitäten sind nach der Festlegung der Faktorenanzahl fixiert.

Der der 1. Ebene zugeordnete Begriff "Population" ist in Klammern gesetzt. Damit soll die Einschränkung dieses Begriffs deutlich gemacht werden. Er gilt in unbeschränktem Umfang nur für die gewählten Populationsfaktorladungen.

Die 5. Ebene dient in der Folge besonders als Möglichkeit, die faktorenanalytischen Ergebnisse daraufhin zu überprüfen, ob die ursprünglich bestimmten Faktoren in der Analyse wieder reproduziert werden, und dabei besonders auf die Reihenfolge in der Eigenwertreihe zu achten.

2. Validitätskoeffizienten aus Simulations- Studien

Der Versuch, TFA und IFA über Simulationsrechnungen zu vergleichen (Schiller und Wittmann 1971), erbrachte keine bedeutsamen Unterschiede der beiden faktorenanalytischen Techniken. Der Grund dafür mag darin liegen, daß dabei größtenteils Aussagen über die Validitätskoeffizienten wie bei Oberla gewonnen werden sollten. Diese Koeffizienten drücken als Korrelation der geschätzten mit den ursprünglichen Faktorwerten nur aus, wieviel Zusammenhang nach einer Analyse noch besteht. Entsprechend war Oberla's Ziel der Nachweis, daß die Faktorenanalyse ihren Anspruch erfüllt. "Der Anspruch der Faktorenanalyse, Größen herauszuarbeiten, die grundsätzlich nicht meßbar sind, die sich aber in meßbaren Variablen ausprägen, ist berechtigt." (Oberla 1968, S.293)

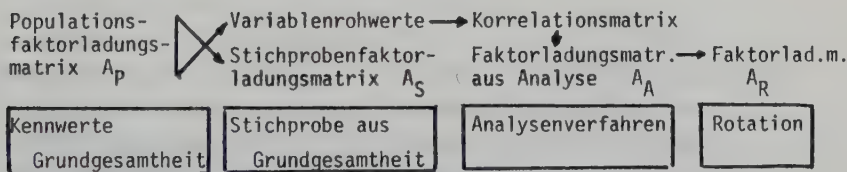
Wenn ein TFA - IFA Vergleich über die Validitätskoeffizienten keine bedeutsamen Unterschiede erbrachte, so heißt das

A. Beide Verfahren sind gleich valide

Diese Aussagen beziehen sich auf das ganze Analysenverfahren, die jeweilige Varimax-Rotation ist mit eingeschlossen. Außerdem ist durch die Bedingung der Einfachstruktur für die Populationsfaktorladungsmatrix der Geltungsbereich eingeschränkt.

Eine vergleichende Studie kann nun nicht auf dieser Basis beendet werden. Es bleibt die Frage, ob nicht andere Wege der Auswertung Unterschiede der faktorenanalytischen Verfahren aufweisen können. Dabei kann unter anderem an die Frage angeknüpft werden, warum die IFA im Bereich mittlerer und hoher Populationsfaktorladungen die 'wahren' Faktoren in der Eigenwertreihe viel deutlicher als die TFA hervorhebt. Diese und andere Fragen können eher über eine intensive Betrachtung der jeweiligen Analysenschritte beantwortet werden. Dazu gehört an dieser Stelle eine Diskussion des Ablaufschemas der Modellrechnung. (s. Oberla 1968, S.274)

Ablaufschema der Modellrechnung



Dieses Ablaufschema ist eine vereinfachte Darstellung des Schemas von Oberla. Verzichtet wurde dabei auf die Hervorhebung der Ebene der Validitätskoeffizientenberechnung, betont wurde dagegen die in den restlichen Ebenen auftretende Möglichkeit des Überprüfens der Auswirkung der einzelnen Schritte.

Ausgangspunkt ist die Kennzeichnung der Grundgesamtheit durch die Populationsfaktorladungsmatrix A_P . Diese bestimmt gleichzeitig die Transformationsmatrix T

$$(1) \quad T = \begin{pmatrix} \sqrt{1-\varphi^2} & 0 \\ \varphi & 1 \end{pmatrix} \quad (\text{Oberla 1978, S.273})$$

φ : Ladung der Variable auf einem Faktor

Diese Transformationsmatrix kann nur für zweidimensionale Faktorenräume mit Einfachstruktur (Variablen laden je nur einen Faktor) Verwendung finden. Eine Erweiterung für mehrdimensionale, auch schiefwinklige Faktorenräume mit der Möglichkeit von Mehrfachladungen haben Schiller u. Wittmann entwickelt. Dabei ergab sich eine Supermatrix T

$$(2) \quad T = \left(\begin{array}{c|c} U_P & B \\ \hline A_P & C \end{array} \right) \quad (\text{Schiller u. Wittmann, 1971, S.72-76})$$

wo U_P : eine Diagonalmatrix mit den Diagonalelementen $\sqrt{1 - \sum g_i^2}$, den Wurzeln aus den Uniqueness-Werten, ist
 A_P : die gewünschte Populationsfaktorladungsmatrix mit den jeweiligen Ladungen der Variablen darstellt
 C : eine Identitätsmatrix (bei orthogonalen Faktoren) ist
 B : eine Nullmatrix ist.

Mit Hilfe eines Zufallszahlengenerators werden n Spaltenvektoren als Variablenrohwerte und k Spaltenvektoren als Faktorrohwerte generiert und jeweils standardisiert. Diese k + n Spaltenvektoren sind, bis auf Zufallseffekte, jeweils orthogonal zueinander. Eine Postmultiplikation der Matrix aus den Spaltenvektoren mit der Transformationsmatrix bewirkt, daß die n untereinander und zu den Faktorwertevektoren senkrechten Variablenwertevektoren so auf den durch die Faktorwertevektoren beschriebenen Raum rotiert werden, daß die gewünschten Ladungen, bis auf zufällige Abweichungen, entstehen.

$$(3) \quad N^Y_n \ N^P_k \times \left(\begin{array}{c|c} n^U_n & n^B_k \\ \hline k^A_n & k^C_k \end{array} \right) = N^Y_n U_n + N^P_k A_n \quad N^Y_n B_k + N^P_k C_k$$

N^Y_n : Matrix der Variablenwerte

N^P_k : Matrix der Faktorwerte

Der erste Teil der Matrix auf der rechten Seite der Gleichung ist die Datenmatrix (Variablen) der Stichprobe. Daraus kann die

Korrelationsmatrix R_S berechnet werden. Außerdem sind die Stichprobenfaktorladungen (A_S als entsprechende Matrix) als Korrelation der Variablen mit den Faktorwerten berechenbar. A_S ist als Schätzung von A_P , R_S als Schätzung von R_P zu betrachten.

R_S , die Stichprobenkorrelationsmatrix, wird einem faktorenanalytischen Verfahren unterworfen. Als Ergebnis kann man erwarten

- A. Eigenwertreihe
- B. Die Faktorladungsmatrix nach der Analyse A_A ,
unter der Voraussetzung der bekannten Anzahl der Faktoren
- C. Die Faktorladungsmatrix A_R nach der Rotation aus A_A ,
ebenfalls unter der Bedingung der bekannten Faktorenanzahl

Die Faktorladungsmatrix A_R kann dazu verwendet werden, die Faktorwerte der Versuchspersonen zu schätzen. Darauf basiert die Technik der Validitätsberechnung als Korrelation der ursprünglichen mit den geschätzten Faktorwerten (Oberla 1968).

Bevor Folgerungen für die hier anzuwendende Auswertungsform gezogen werden, soll noch ein kritischer Blick auf Oberla's Auswertung seiner Ergebnisse folgen.

Oberla stellt größtenteils seine Validitätskoeffizienten in Abhängigkeit von der mittleren Populationsfaktorladung dar. Das führt auch zu der Darstellung für gleiche mittlere Ladungen bei schwacher bzw. starker Streuung. (Oberla 1968, S.283)

An dieser Stelle muß aber die Interpretation der Ergebnisse zu falschen Schlüssen führen. "Die Validitäten der stärker streuenden Serie liegen deutlich höher als die der weniger stark streuenden Serie - bei gleicher durchschnittlicher Beziehung zwischen Variablen und Faktoren." (Oberla S.283)

Liegen hier wirklich 'gleiche durchschnittliche Beziehungen' vor? Tatsächlich sind die mittleren Populationsladungen gleich hoch. Diese mittleren Ladungen sind aber nicht die Kennwerte, die in einer Analyse dann die Ausprägung der Faktoren bestimmen. Vielmehr muß hier von den Quadraten der Ladungen, den Kommunalitäten,

Varianzbeiträge der Variablen bestimmt - und in Abhängigkeit von diesen Varianzbeiträgen wird sich auch die Höhe des Validitätskoeffizienten ausprägen.

Beispiel: Einfachladungen in der Population

Variable	Faktorladungen auf dem 1.Faktor		Kommunalitäten	
	Streuung		Streuung	
	Schwach	Stark	Schwach	Stark
1	.40	.70	.16	.49
2	- .40	- .55	.16	.30
3	.40	.40	.16	.16
4	- .40	- .25	.16	.06
5	.40	.10	.16	.01
Mittl.Betrag	.40	.40	.16	.20

Dieses einfache Beispiel zeigt, daß die stärker streuende Serie in Wirklichkeit im Vergleich zu der schwach streuenden Serie 25% mehr Varianz einbringt. Die unterschiedlich hohen Validitäten sind also nicht auf die Streuungen der Ladungen, sondern auf die unterschiedlich hohen Varianzbeiträge zurückzuführen.

Folgerungen für die vorliegende Untersuchung

Die einzelnen Ebenen (Schritte) der Analysen müssen stärker in Betracht gezogen werden.

Die Auswertung und Interpretation der Ergebnisse muß im Hinblick auf die Populations- bzw. Stichprobenkommunalitäten durchgeführt werden.

Oberla erhält auch bei niedrigen Ladungen relativ hohe Validitäten. Diese ließen sich bei Schiller und Wittmann nicht wiederholen, solange man 'blind', d.h. ohne Vorkenntnisse, die gesamte Analyse durchführte. Fehlerfaktoren traten dann in der Eigenwertreihe vor die eigentlichen Faktoren.

3. Fehlerfaktoren

Die vorliegende Arbeit kann als 'reine' Monte-Carlo-Studie betrachtet werden. Dasselbe gilt für Oberla's Simulationen. Der Fehler, der hier den bedeutsamsten Stellenwert erhält, ist danach (Winer 1968)

- a. Stichprobenfehler
- b. Fehler, die auf Grund mangelnder Anpassung eines Modells an empirisch gewonnene Daten entstehen

Setzen wir hier statt 'empirisch gewonnene' 'simulierte' Daten, dann werden beide Fehlerarten zu berücksichtigen sein.

Vorläufig sollen nur die Stichprobenfehler interessieren. Der 'Anpassungsfehler' wird nur in Abhängigkeit von dem Stichprobenfehler darstellbar werden. Bei dem Vergleich verschiedener faktorenanalytischer Verfahren muß er sich allerdings im Verhältnis der Güte der Reproduktion der simulierten Strukturen erkennbar auswirken.

Stichprobenfehler bedeutet hier: Stichproben sind in Abhängigkeit von ihren Kennwerten - Versuchspersonenanzahl N , Variablenanzahl n , Faktorenanzahl k , Höhe der Kommunalitäten - durch zufällige Abweichungen vom Populationswert gekennzeichnet. Für den ersten Schritt bei der Simulation heißt das: Die $k + n$ generierten Vektoren sind nicht völlig unabhängig voneinander, sie weisen je nach der Kombination der Kennwerte der Stichprobe zufällige Interkorrelationen auf. Im Hinblick auf die Faktorenanalyse bedeutet dies die Erwartung eines systematischen Fehlers, eines Bias. Da für alle möglichen Stichproben immer eine zufällige gemeinsame Varianz zu erwarten ist, kann kein faktorenanalytisches Ergebnis eine 'unbiased' Schätzung der Populationsparameter liefern. Eine augenfällige Demonstration dieser Tatsache sind die Eigenwertreihen hauptkomponentenanalysierter Korrelationsmatrizen aus Zufallszahlen (Horn 1965, S.180 u. S.184' / Oberla 1968, S.127) Es ergeben sich in Abhängigkeit von der Variablenanzahl eine Reihe von Eigenwerten größer 1. Horn schlägt im Hinblick auf diesen Bias vor, das Problem der Anzahl der bedeutsamen Faktoren folgendermaßen zu lösen: Es wird entsprechend der zu analysierenden

Korrelationsmatrix eine Matrix aus Zufallszahlen bei gleicher Versuchspersonen- und Variablenanzahl berechnet. Die Anzahl interpretierbarer Faktoren ist dann bestimmbar, wenn den Eigenwerten der eigentlichen Analyse jeweils der aus der Zufallsmatrix geschätzte Betrag der zufälligen gemeinsamen Varianz abgezogen wird. Mit anderen Worten heißt das, daß jeder Eigenwert um den geschätzten Bias vermindert wird. Eigenwerte, die danach noch größer als 1 sind, können als interpretierbare Faktoren angesprochen werden.

Horns Vorschlag basiert auf der Annahme, daß sich ein Bias additiv in den Eigenwerten auswirkt. Diese Annahme soll hier übernommen werden. Kritisch zu sehen sind zwei weitere nicht berücksichtigte Punkte. Der erste ist das Problem von Fehlerfaktoren, von Faktoren nur zufälliger gemeinsamer Varianz (s. R.B.Cattell, 1966a, S.202). Es ist durchaus vorstellbar, daß solch ein Fehlerfaktor das Gewicht erhält, das ihn zwischen oder vor die eigentlichen Faktoren einreicht. Der zweite Punkt ist die nicht berücksichtigte Abhängigkeit des Bias von den Kommunalitäten bzw. Uniqueness-Anteilen der Variablen. Für h^2 gegen 1 bzw. u^2 gegen 0 muß man erwarten, daß der Betrag des Bias gegen Null strebt. In diesem Falle wäre die Anwendung der Horn'schen Empfehlung nicht gerechtfertigt.

An dieser Stelle muß noch einmal das Problem der vor den eigentlichen Faktoren auftretenden Fehlerfaktoren betrachtet werden. Es mußte schon die Frage gestellt werden, wie Oberla bei niedrigen Ladungen noch die relativ hohen Validitätskoeffizienten erhalten konnte. In der Studie von Schiller und Wittmann waren diese Validitäten nicht ohne eine besondere Technik zu gewinnen. Diese kann als Identifikationstechnik zur Bestimmung der ursprünglichen Faktoren gekennzeichnet werden.

Die k generierten Faktorwertvektoren werden dabei nicht nur mit den ersten k nach der Analyse geschätzten Faktorwertvektoren, sondern mit den ersten $k+2$ geschätzten Faktorwertvektoren korreliert. Dabei gilt die Annahme, daß die simulierten Faktorwerte nicht mehr als zufällig mit den geschätzten Faktorwerten korreliert sind. Ein ursprünglicher Faktor wird sich stärker bemerkbar machen. Auf diesem Weg ist feststellbar, bei welchen Populationskennwerten

bzw. welchem Variablen- und Versuchspersonen-'sampling' Fehlerfaktoren in dem Analysenergebnis zu erwarten sind.

Beispiel für nach der Rotation erhaltene Validitätskoeffizienten

$N = 100$, $n = 10$, $k = 2$, Einfachladungen .30, je Faktor

5 Variablen

1) Stichprobe Nr.2

TFA1	<u>.429</u>	-.380	.125	.001	.200	.196	.164	<u>.516</u>
IFA	<u>.599</u>	-.147	.139	-.263	.199	<u>.307</u>	-.013	.146

2) Stichprobe Nr.3

TFA1	<u>-.524</u>	-.123	.032	.005	.255	<u>-.320</u>	.156	<u>.312</u>
IFA	<u>-.401</u>	<u>-.249</u>	.039	.130	<u>.318</u>	-.093	.211	.044

Interpretation :

1) TFA: Der ursprünglich erste Faktor wird durch die beiden ersten Faktoren nach der Analyse beschrieben, der ursprünglich zweite durch den vierten Faktor nach der Analyse.

IFA: Ursprüngliche und nach der Analyse erhaltene Faktoren stimmen überein.

2) TFA: Ursprüngliche und nach der Analyse erhaltene Faktoren stimmen überein.

IFA: Zuordnung ist schwierig. Die beiden ersten Faktoren nach der Analyse beschreiben den ursprünglich ersten, der erste nach der Analyse auch den ursprünglich zweiten Faktor.

Für beide Stichproben ergaben sich aus der Hauptkomponentenanalyse vier Eigenwerte größer 1! Zumindest für die TFA1 der ersten aufgeführten Stichprobe kann ein an 3.Position auftretender Fehlerfaktor festgestellt werden, wobei der zweite zu erwartende Faktor erst die Position des 4. Faktors einnimmt.

Folgerungen

Horns Vorschlag ist in der Praxis kaum durchführbar. Es müßte aber

im Rahmen dieser Untersuchung möglich sein, den Betrag der zusätzlichen zufälligen Varianz für den Praktiker bestimmbar zu machen. Das Hauptproblem wird dabei die Erarbeitung unabhängig von der Vorkenntnis der Anzahl der Faktoren sein.

Überla berichtet von seinen Validitätskoeffizienten auch im Bereich niedriger Populationsfaktorladungen, ohne dabei das Problem der Fehlerfaktoren zu berücksichtigen. Die oben aufgeführten Beispiele zeigen, daß es möglich sein wird, kritische Bereiche zu beschreiben, wo Fehlerfaktoren mit großer Wahrscheinlichkeit zu erwarten sind. Der kritische Bereich wird von den Faktorladungen in der Population, der Variablenanzahl und der Versuchspersonenanzahl abhängen. Er muß als Grenze für sichere Lösungen abhängig von den genannten Kennwerten von dem Praktiker bestimmbar sein.

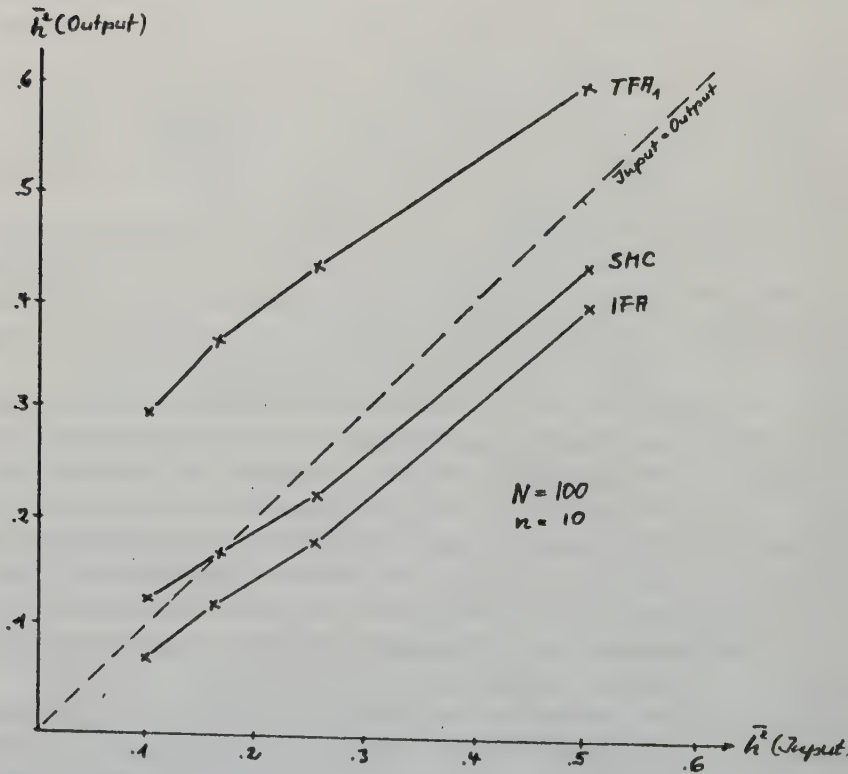
Der 'Anpassungsfehler' des Modells an die Daten wird kaum unabhängig vom Stichprobenfehler darstellbar sein. Auf der anderen Seite soll es Ziel dieser Studie sein, den Umfang und die Genauigkeit der Anpassung der verschiedenen Analysenverfahren zu beschreiben. Darüberhinaus wäre es wünschenswert, falls möglich, Techniken zu erarbeiten und zu berichten, die es ermöglichen, die Anpassung im Hinblick auf größtmögliche Genauigkeit zu steuern. Dabei ist unter anderem an die Bestimmung der Faktorenanzahl zu denken.

4. Multiple Korrelationskoeffizienten und die Kommunalitätsschätzung

Mit den multiplen Korrelationen und den daraus resultierenden SMCs (Squared Multiple Corr.) als Kommunalitäten-Schätzmöglichkeit ergeben sich Werte, die sowohl der Ebene der Simulation wie der Analyse zuzuordnen sind. Sie können als Kennwerte der Stichprobe betrachtet oder als Analysenergebnisse mit Schätzfunktion in Bezug auf die Population interpretiert werden. Durch diese Besonderheit gewinnen die SMCs eine zentrale Bedeutung für diese Simulationsstudie. Dies wird durch weitere Kennzeichen der SMCs noch betont:

- 1) Die SMCs sind eindeutige Werte, berechenbar für jede Korrelationsmatrix
- 2) Als gemeinsame Varianz einer Variablen mit allen restlichen Variablen in einer Stichprobe ist zu erwarten, daß die SMC weniger durch zufällige gemeinsame Varianz zwischen zwei Variablen beeinflußt wird als die einfachen Korrelationen
- 3) Jede einzelne SMC wie auch ein Mittel aus den SMCs einer Stichprobe kann als Schätzung des Varianzanteiles am Raum der gemeinsamen Varianz betrachtet werden
- 4) Damit gewinnt der Begriff der 'beobachteten Kommunalitäten' von Wrigley an Bedeutung. Für diese Studie erscheint besonders wichtig, daß die SMCs keine Vorentscheidungen über die Anzahl der Faktoren erwarten, aber dennoch Aussagen über den Varianzanteil des Raums der gemeinsamen Varianz zulassen
- 5) Die SMCs werden im allgemeinen niedriger sein als die Kommunalitäten aus einer Analyse (Guttman's "lower bound"), Letztere stellen die gemeinsame Varianz einer Variablen mit den Faktoren dar. Die Differenz zwischen den Kommunalitäten der Stichprobe und den SMCs könnte als systematischer Fehler oder Bias bei der Schätzung der Kommunalitäten durch die SMCs beschrieben werden. Diese Differenz, dieser Bias müßte in seiner Abhängigkeit von den Kennwerten der Stichprobe - Versuchspersonen- und Variablenanzahl, Kommunalitäten - bestimmbar sein.
- 6) Die Darstellung 1 auf der nächsten Seite ist eine erste Datenanalyse unter den beschriebenen Gesichtspunkten. Die gestrichelte Linie stellt die erwarteten Kommunalitäten dar. Zusätzlich zu den mittleren SMC-Werten sind die mittleren Kommunalitäten aus der Hauptkomponentenanalyse und der Image-Analyse vorgestellt. Eine erste Interpretation zeigt, daß der eingeschlagene Weg der Vergleiche der einzelnen Analyseverfahren über die mittleren Kommunalitäten durchaus gangbar ist. Außerdem zeigt sich, daß die mittleren SMCs sich ohne Vorkenntnis der Faktorenanzahl den beiden Analyseverfahren mit dieser Vorkenntnis bei der Schätzung der gemeinsamen Varianz eher überlegen zeigen.

DARSTELLUNG 1 : Mittlere Kommunalitäten bei Input (Stichprobe)
und Output (Nach der Analyse)



Folgerungen für die vorliegende Untersuchung

Die SMCs sollen zu einer 'Führungsgröße' bei der Erarbeitung von Schätzfunktionen für die 'wahre' gemeinsame Varianz, den Varianzanteil des Faktorenraumes, erhoben werden.

Die Bestimmung des Varianzanteiles des Faktorenraumes ist für den Praktiker normalerweise an die Bestimmung der Anzahl der Faktoren gebunden. Dabei geht es darum, die Eigenwertreihe daraufhin zu überprüfen, welche Anzahl von Faktoren für die jeweiligen Analysen-

ergebnisse adäquat erscheint. Kennzeichnend für die Vielzahl möglicher Wege und Kombinationen dieser Wege zur Faktorenanzahlbestimmung (z.B. Fürntratt) ist, daß sie keine 'blinde', keine vom jeweiligen Forscher unabhängige Durchführung der Faktorenanalyse gestatten. Über die SMCs ist es eventuell möglich, eine Analyse blind durchzuführen.

Eine Konsequenz der Orientierung an den SMCs und entsprechend an den jeweiligen Kommunalitäten zum Vergleich der Ebenen ist, daß das Rotationsproblem hier unberührt bleiben muß. Die Rotationen verändern den Betrag der Kommunalitäten der Variablen nicht.

Bei der vergleichenden Betrachtung der Kommunalitäten der Population (aus der Populationsfaktorladungsmatrix) und der Stichprobe hat sich gezeigt, daß die Simulationen erwartungstreue Schätzungen der Populationskennwerte liefern. Für die weitere Untersuchung bedeutet das den Verzicht auf vergleichende Betrachtungen mit den Populationskennwerten. Wenn in der Stichprobe zufällige Abweichungen von der gewünschten (Populations-)Struktur auftreten, dann gilt die Forderung, daß eine Faktorenanalyse die jeweilige Faktorenstruktur der Stichprobe reproduzieren können muß. Oberla (1968, S.278) stellt für sein sehr ähnliches Simulationsprogramm ebenfalls fest, daß tatsächlich Stichproben aus der gewünschten Grundgesamtheit gewonnen werden. Er entscheidet sich, unter dem Gesichtspunkt der Fragestellung bei seiner Simulation, für die mittleren Korrelationen der Variablen in der Stichprobenkorrelationsmatrix als unabhängige Variable bei der Darstellung der Validitätskoeffizienten.

5. Measure of Sampling Adequacy (MSA)

Die Zielsetzung dieser Arbeit entstand unter anderem in Anlehnung an die Hinweise Guttman's (1953), die Beurteilung von Daten einer Stichprobe über R^{-1} , die Inverse der Korrelationsmatrix, bzw. die Anti-Image-Kovarianzmatrix Q durchzuführen. Diesen Hinweis Guttman's verwendete Kaiser (1970, 1974) bei der Erarbeitung seiner MSA- Formeln.

Vor eine Betrachtung der MSA-Koeffizienten soll eine kurze Betrachtung der Image-Theorie gestellt werden. Sie dient zur Klärung der Aussagen eines MSA- Koeffizienten und zur Interpretation der Ergebnisse aus der Image-Analyse.

5.1. Exkurs zur Image - Theorie

Guttman's Image - Theorie (1953) basiert auf der multiplen Regressionsrechnung. Jeder (standardisierte) Variablenvektor eines Datensatzes wird in einen Image-Vektor und einen Anti-Image-Vektor zerlegt. Das Image stellt einen geschätzten Wert des Anteils am Raum der gemeinsamen Varianz dar, eine Vorhersage aus den restlichen $n-1$ Variablen:

$$(4) \quad p_j = \sum_{i=1}^n w_{ji} z_i \quad w_{jj} = 0$$

Das Image p_j als geschätzter Vektor aus der multiplen Regression ist eine Linearkombination aus den restlichen Variablen. Die w_{ji} sind die jeweiligen Regressionskoeffizienten. Der Anti-Image-Vektor e_j kann als Schätzfehler gekennzeichnet werden. Hier sind die Anteile einer Variablen erfaßt, die nicht durch die restlichen Variablen schätzbar sind:

$$(5) \quad z_j = p_j + e_j$$

Nach Guttman kann nun die Korrelation zwischen zwei standardisierten Variablen eines Variablensatzes von n Variablen gezeigt werden als Differenz der Kovarianzen der entsprechenden Images bzw. Anti-Images :

$$(6) \quad r_{ji} = g_{ji} - \delta_{ji}$$

$$j \neq i$$

g_{ji} ist die Kovarianz der Images

δ_{ji}^* ist die Kovarianz der Anti-Images

Außerdem beweist Guttman, daß die Kovarianz der Anti-Images zweier Variablen gleich dem negativen Kovarianzbetrag ist, der nicht durch die restlichen Variablen erklärt werden kann. Es handelt sich hier um Kovarianzen außerhalb des durch die Images beschriebenen Raumes der gemeinsamen Varianzen.

In Matrizenschreibweise läßt sich dann die Image- Kovarianzmatrix darstellen als

$$(7) \quad G = R + Q^* - S^2$$

R : Korrelationsmatrix

Q_2 : Anti-Image-Kovarianzmatrix mit Nullen in der Diagonalen

S^2 : Diagonalmatrix der Anti-Image-Varianzen

Bisher wurde im Sinne Gutmans von den 'Partial Images' als Images gesprochen. Die Beziehung der Image - Theorie zur Theorie der gemeinsamen Faktoren in der Faktorenanalyse zeigt sich beim Übergang von der Stichprobenebene zum 'Universe of Content'. Die 'Partial Images' werden zu 'Total Images', und nach Guttman (1956) gilt unter der Bedingung

$$\lim_{n \rightarrow \infty} \frac{k}{n} = 0$$

k : Anzahl der gemeins.
Faktoren
 n : Variablenanzahl

daß

Q^* zu einer Nullmatrix wird, es keine Kovarianzen außerhalb des Raumes der gemeinsamen Faktoren gibt

S^2 in die Diagonalmatrix der Uniqueness- Anteile übergeht.

Damit wird die Image- Kovarianzmatrix G zur reduzierten Korrelationsmatrix $R - U^2$. Für jede endliche Variablenanzahl ist Q keine Nullmatrix und S^2 setzt sich aus den SMCs der Variablen zusammen. Beides bewirkt bei einer Faktorenanalyse der Image-Kovarianzmatrix eine Unterschätzung der Anteile an der gemeinsamen Varianz.

5.2. Die Beurteilung der Stichprobenadäquatheit

Kaiser (1970 u. 1974) benutzt die Anti-Image-Korrelationsmatrix, um im Sinne von Guttman's Hinweis auf die Bedeutung dieser Matrix für die Beurteilung einer Stichprobe einen Koeffizienten für solch eine Beurteilung aus dieser Matrix zu berechnen. Dieses "Measure of Sampling Adequacy" hat folgende Form

$$(8) \quad MSA_{\text{Stichprobe}} = 1 - \frac{\sum_{i=1}^n \sum_{j=1}^n q_{ij}^2}{\sum_{i=1}^n \sum_{j=1}^n r_{ij}^2} \quad i \neq j \quad (\text{Kaiser 1970})$$

umgeformt

$$(9) \quad = \frac{\sum_{i=1}^n \sum_{j=1}^n r_{ij}^2 - \sum_{i=1}^n \sum_{j=1}^n q_{ij}^2}{\sum_{i=1}^n \sum_{j=1}^n r_{ij}^2} \quad i \neq j$$

$$(10) \quad MSA_{\text{Stichprobe}} = \frac{\sum_{i=1}^n \sum_{j=1}^n r_{ij}^2}{\sum_{i=1}^n \sum_{j=1}^n r_{ij}^2 + \sum_{i=1}^n \sum_{j=1}^n q_{ij}^2} \quad i \neq j \quad (\text{Kaiser 1974})$$

Die Formel (8) bzw. (9) kann in Gleichung (10) überführt werden, indem bei (9) in Zähler und Nenner die Summe der nichtdiagonalen quadr. Werte der Anti-Image-Korrelationsmatrix addiert wird. Dadurch wird erreicht, daß die neuen MSA nicht mehr negativ werden können.

$\sum \sum q_{ij}^2$ bedeutet eine Aufsummierung der 'Dublettenvarianz', der nur jeweils zwei Variablen gemeinsamen Varianz. Diese ist im Ansatz der multiplen Regression nicht durch die restlichen $n-2$ Variablen bestimmbar.

Bildet man

$$\frac{\sum_{i=1}^n \sum_{j=1}^n q_{ij}^2}{2} \quad i \neq j$$

so liegt eine Schätzung des gemeinsamen Varianzanteils eines Variablensatzes, der jeweils nur zwei Variablen gemeinsam ist, vor. Dabei kann es nach den Voraussetzungen keine Überschneidungen von den einzelnen Varianzanteilen geben.

Bildet man entsprechend

$$\frac{\sum_{i=1}^n \sum_{j=1}^n r_{ij}^2}{2} \quad i \neq j$$

so sind viele Varianzüberschneidungen möglich. An und für sich gleiche Varianzanteile werden mehrfach summiert. Das Ausmaß der Mehrfachsummation hängt wesentlich von der Anzahl der Variablen je Faktor ab.

Betrachtet man nun das von Kaiser bei seinen MSA-Koeffizienten ausgenutzte Verhältnis dieser beiden Summen, so muß man sich fragen, ob die scheinbare Einfachheit dieses Verhältnisses die gewünschte Aussagekraft erhalten kann. Im Sinne der früher aufgestellten Forderung, möglichst einzelne Ebenen zu betrachten, könnte man hier die Ebene der Image - Erarbeitung (Ebene der multiplen Regression bzw. Korrelation) beibehalten. Die Konsequenz dieser kritischen Anmerkungen wäre ein MSA, bei dem $\sum \sum r_{ij}^2$ durch die SMCs ersetzt würde:

$$(11) \quad MSA_{\text{Stichprobe}} = \frac{\sum_{i=1}^n SMC_i}{\sum_{i=1}^n SMC_i + \sum_{i=1}^n \sum_{j=1}^n q_{ij}^2} \quad i \neq j$$

Hier ist die geschätzte gemeinsame Varianz eines Variablensatzes

ins Verhältnis gesetzt zu der Summe aus dieser Varianz und der Dublettenvarianz. Man kann zu diesem Verhältnis die Hypothese erstellen, daß bei Werten unter .50 das Sampling nicht mehr adäquat ist. Der Betrag der Dublettenvarianz übersteigt den Betrag der geschätzten gemeinsamen Varianz. Fehlerfaktoren können sich in der Analyse als eigentliche Faktoren darstellen.

Nach Wittmann (1976) gibt es noch einige Probleme bei der Beurteilung der MSA. Derzeit steht kein Signifikanztest zur Verfügung, letzten Endes ist die Verteilungsfunktion des MSA noch nicht bekannt. Kaiser (1970) gibt nur Klassifizierungshinweise. Dabei ist die Übereinstimmung von MSA und Professor Tuckers subjektiver Einstufung der Qualität der Daten bei 54 Korrelationsmatrizen in Höhe von $r = .85$ bedeutsam (Kaiser 1970). Es ist zu fragen, welche Information Tucker verwendet hat, um diese hohe Übereinstimmung zu erzielen. Betrachtet man dazu Tuckers Aufsatz mit den Daten zu diesen Korrelationsmatrizen (Tucker, Koopman, Linn 1969), dann läßt sich vermuten, daß Tuckers hauptsächliches Kriterium in den SMCs zu sehen ist. Läßt sich der MSA durch eine einfache Betrachtung der SMCs bzw. einem mittleren Wert aus diesen ersetzen?

Zu dieser Frage können auch die MSA auf der Ebene der einzelnen Variablen herangezogen werden. Diese sind nach Kaiser ein Maß für den Grad des 'belonging to the family'. Einen entsprechenden Aussagegehalt kann auch den SMCs zugeschrieben werden. In ihrer Funktion als Schätzung der Kommunalitäten der einzelnen Variablen sollen sie eine Aussage über den ihnen eigenen Varianzanteil am Raum der gemeinsamen Varianz liefern. Der Umfang eines 'belonging to the family' wird geschätzt.

6. Zusammenfassung und erste Zielsetzung

Die Möglichkeit in der Monte-Carlo-Studie, ausgehend von den bekannten Populationskennwerten bis hin zu den Stichprobenergebnissen aus einer Faktorenanalyse, jeden einzelnen Schritt, jede einzelne Ebene verfolgen zu können, soll ausgenutzt werden. Dabei interessiert über die reine Prüfung der Angemessenheit der Daten für eine Analyse (MSA von Kaiser (1970) oder 'Test for sphericity' (Bartlett nach Cooley u. Lohnes (1971)) hinaus der Umfang eines systematischen Fehlers oder Bias in Abhängigkeit von den vier 'main effects' (Kaiser 1970):

- A. Variablenanzahl n
- B. Versuchspersonenanzahl N
- C. Faktorenanzahl k
- D. Durchschnittliche Höhe der Interkorrelationen $\bar{r}$

Die Faktorenanzahl als unabhängige Größe soll, soweit möglich, ausgeklammert werden. Wie oben begründet, läßt sich das über die SMCs als Kommunalitätenschätzungen durchführen. Dahinter steht die Absicht, die von der Faktorenanzahl nicht abhängigen SMCs (die den $\bar{r}$ - Wert hier ersetzen werden) zur Bestimmung der Faktorenanzahl bei bekanntem n und N zu nutzen.

Durch die vorliegenden Daten der Population können die $\bar{r}$ einerseits durch die Populationsfaktorladungen, andererseits auf der Ebene der Stichprobe durch die Wurzel aus den SMCs ersetzt werden. Wie schon weiter oben begründet, werden hier nicht die Korrelationen, sondern die Kommunalitäten bzw. die ihnen entsprechenden SMCs für eine Analyse der simulierten Daten verwendet.

Damit kann man die Problemstellung folgendermaßen kennzeichnen, wobei im folgenden meist nur noch die Stichprobenkennwerte berücksichtigt werden:

Es soll die Faktorladungsmatrix aus einer Analyse gekennzeichnet und bestimmt werden als Funktion der Faktorladungsmatrix der

Stichprobe, der Versuchspersonen- und der Variablenanzahl. Statt der Ladungsmatrizen die jeweiligen mittleren Kommunalitäten eingesetzt, ergibt sich folgende zu erarbeitende Funktion:

$$(12) \quad \bar{h}_A^2 = f(\bar{h}_S^2; N; n)$$

wobei bedeutet

$\bar{h}_S^2$: mittlere Kommunalität der Stichprobe

$\bar{h}_A^2$: mittlere Kommunalität der jeweiligen Analyse

Auf der Ebene der einzelnen Variablen ergibt sich die entsprechende zu erarbeitende Funktion

$$(13) \quad h_A^2 = f(h_S^2; N; n), \text{ wobei hier die jeweiligen Kommunalitäten der einzelnen Variablen angesprochen sind.}$$

Wenn sich diese Funktionen bestimmen lassen, dann können sie für die jeweiligen faktorenanalytischen Verfahren angeben, wie genau die 'wahre' Struktur der Daten in dem Analysenergebnis erscheint. Im Hinblick auf das Eigenwertproblem kann eine Lösung in der Form erwartet werden, daß solch eine Funktion die irrelevanten Varianzanteile zu bestimmen weiß.

In Anlehnung an die Darstellung 1 (S.14) lassen sich die Formeln 12 und 13 auch in Form von Hypothesen schreiben als

$$(14) \quad \begin{aligned} \bar{h}_A^2 &= \bar{h}_S^2 + K \bar{h}_S^2 \\ &= \bar{h}_S^2 (1 + K) \end{aligned}$$

$$\text{wobei } K = f(N; n)$$

$$(15) \quad h_A^2 = h_S^2 (1 + K)$$

Die funktionale Abhängigkeit der Größe K von N und n zeigt hier verdichtet das Problem der Koppelung von 'Personensampling' und 'Variablensampling'.

Auf der Ebene der einzelnen Variablen bedeutet die Abhängigkeit von h_S^2 : Wie weit werden niedrige Kommunalitäten in der Stichprobe auch durch die jeweilige Faktorenanalyse wieder als niedrige Kommunalitäten reproduziert, wie weit werden hohe Kommunalitäten wieder als solche aufgezeigt? Die Kennzeichnungen als 'hoch' bzw. 'niedrig' werden sich an der durchschnittlichen Kommunalität einer Stichprobe orientieren müssen.

III. Betrachtungen zum Grundkonzept der Faktorenanalyse

1. Die Aufspaltung der Einheitsvarianz

An dieser Stelle soll noch einmal von den Populationswerten ausgegangen werden und mögliche Konsequenzen des 'Sampling' betrachtet werden. Die Aufspaltung der Einheitsvarianz einer Variablen läßt sich fassen als

$$\begin{aligned} s^2 &= 1.00 \quad (\text{das entspricht 100\% Varianz}) \\ &= \text{Kommunalität } h^2 \text{ plus Uniqueness } u^2 \\ &= \text{Summe der Faktorladungsquadrate } a_1^2 \text{ plus Uniqueness} \end{aligned}$$

Wenn diese Werte für eine Population festgelegt sind, dann läßt sich für ein Versuchspersonensampling erwarten, daß die Stichprobenwerte jeweils um den Populationswert streuen, und zwar in Abhängigkeit von der Versuchspersonenanzahl.

Für den Bereich des Variablensampling gilt für die Population: Kommunalität und Uniqueness sind unabhängig voneinander für jede einzelne Variable, außerdem zeigen die Uniqueness-Anteile der verschiedenen Variablen keine gemeinsamen Anteile. Für jede Stichprobe läßt sich nun erwarten, daß diese Bedingungen der Null-Kovarianzen nicht erfüllt sind. Es wird zufällige gemeinsame Varianz geben, die vor allem als gemeinsame Varianz der unquenen Anteile der Variablen die Ergebnisse einer Faktorenanalyse bedeutsam verfälschen kann. Im gleichen Umfang sind auch zufällige gemeinsame Varianzanteile der unquenen Anteile der Variablen mit den jeweiligen Anteilen am Raum der gemeinsamen Varianz der restlichen Variablen zu erwarten.

Für die Image-Analyse läßt sich daraus zunächst der Schluß ziehen, daß solche Kovarianzen über die Anti-Image-Kovarianzmatrix der Analyse entzogen werden, während bei einer Hauptkomponentenanalyse diese sich bemerkbar machen müssen.

2. Die Bedingungen der Null - Kovarianzen

Die in der Transformationsmatrix festgelegten Varianzaufteilungen für die einzelnen Variablen können als Beschreibungen dieses bestimmten Teils des 'universe of content' gesehen werden. Auf der Basis dieser Populationskennwerte wird mit der Simulation eine Stichprobe gezogen. Die dazugehörige Korrelationsmatrix wird faktorenanalysiert. Differenziert man nun vorläufig nur zwischen der Image- und der Hauptkomponentenanalyse, dann ergibt sich für die zu erwartenden Null-Kovarianzen folgendes Bild:

IFA

$$\left. \begin{array}{l} \delta_{1j}^{un} = E \left[\begin{matrix} e_1^{(n)} & e_j^{(n)} \end{matrix} \right] \longrightarrow E \left[\begin{matrix} e_1^{(\infty)} & e_j^{(\infty)} \end{matrix} \right] \\ E \left[\begin{matrix} u_1^{(n)} & u_j^{(n)} \end{matrix} \right] = 0 \end{array} \right\} = E \left[\begin{matrix} u_1^{(\infty)} & u_j^{(\infty)} \end{matrix} \right] = 0$$

Kennzeichnung für die IFA :

Für die hier in Betracht gezogene TFA als Hauptkomponentenanalyse gilt, daß sie zu erwartende Kovarianzen der unigen Anteile der Variablen über die partielle Korrelation ausschöpft und als Varianz des Raumes der gemeinsamen Varianz zu beschreiben versucht. Für die Analyse als Verfahren gilt die Bedingung

$$E \left[\begin{matrix} u_i & u_j \end{matrix} \right] = 0$$

Die unigen Anteile der Variablen i und j sollen keine Kovarianz aufweisen. Auf der anderen Seite ist für jede Stichprobe im Vergleich zu den Populationswerten zu erwarten, daß die unigen Anteile zweier Variablen Kovarianz aufweisen, das heißt

$$h_{\text{Population}}^2 \leq h_{\text{TFA}}^2$$

Dies kommt in seiner Bedeutung einem auf jeden Fall zu erwartenden systematischen Fehler gleich. Für diesen Fehler können noch weitere Folgerungen beschrieben werden. Dabei geht es um die

Abhängigkeit dieses systematischen Fehlers von der Höhe der jeweiligen unigen Anteile der Variablen.

- A. Wird der unique Anteil einer Variablen sehr gering, so ist die Erwartung von Kovarianzen mit den unigen Teilen anderer Variablen ebenfalls gering.

Für $u_i^{(n)} \rightarrow 0$ ist

$E [u_i^{(n)} u_j^{(n)}] \rightarrow 0$ zu erwarten.

- B. Je größer die unigen Anteile zweier Variablen sind, um so größer wird der Betrag der Kovarianzen dieser Anteile zu erwarten sein.

Für $u_i^{(n)} \rightarrow 1$ und $u_j^{(n)} \rightarrow 1$ ist

$E [u_i^{(n)} u_j^{(n)}] \rightarrow \text{Maximum}$ zu erwarten.

Kennzeichnung der IFA

Da die δ_{ij}^n als Kovarianzen der Anti-Images der Analyse entzogen werden und nur die Image-Kovarianzmatrix faktorisiert wird, kann man für die Schätzungen der Kommunalitäten der Stichprobe aus den Analysenergebnissen erwarten, daß sie bias-frei sind.

Diese Betrachtungen zu den Kovarianzen der unigen Anteile der Variablen gelten entsprechend für mögliche Kovarianzen der unigen Teile der einen mit den 'gemeinsamen' Teilen der anderen Variablen. Hier gelten entsprechende Bedingungen von Null-Kovarianzen:

IFA

$$\left. \begin{array}{l} E \left[\begin{pmatrix} p_1^{(n)} & p_j^{(n)} \end{pmatrix} \right] \longrightarrow E \left[\begin{pmatrix} p_1^{(\infty)} & p_j^{(\infty)} \end{pmatrix} \right] \\ E \left[\begin{pmatrix} c_1^{(n)} & c_j^{(n)} \end{pmatrix} \right] \end{array} \right\} = E \left[\begin{pmatrix} c_1^{(\infty)} & c_j^{(\infty)} \end{pmatrix} \right] = 0$$

TFA

Auch hier werden für Stichproben zusätzliche Kovarianzen erwartet werden müssen. Die IFA wird diese ebenfalls als gemeinsame Varianz beschreiben, allerdings durch den multiplen Regressionsansatz in geringerem Umfang, als es bei der TFA der Fall sein wird.

Diese Betrachtungen zu einem möglichen Bias erfolgten nur bezogen auf die Image- und die Hauptkomponentenanalyse. Die weiteren, in die vorliegende Untersuchung einbezogenen Analysenverfahren, verwenden die SMCs als Kommunalitätenschätzungen. Gibt es bei diesen SMCs einen Bias, dann wird sich dieser entsprechend auswirken.

Da die SMCs hier im Vordergrund der Betrachtungen bleiben sollen, muß auch für sie der mögliche Bias in Abhängigkeit von den Stichprobenkennwerten in Erwägung gezogen werden. Dabei können folgende Erwartungen beschrieben werden:

- A. Bei durchschnittlich niedrigen Kommunalitäten werden die SMCs ebenfalls niedrig sein. Zufällige Kovarianzen der oben beschriebenen Art können einen geringen positiven Bias bewirken.
- B. Bei durchschnittlich hohen Kommunalitäten wird die Varianz der Variablen durch die Faktoren erschöpfend beschrieben. Die SMCs werden ebenfalls hohe Werte liefern, da die Vorhersage einer Variablen aus den restlichen Variablen fast vollständig gelingt.
- C. Für durchschnittlich mittlere Kommunalitäten werden die SMCs zu niedrige Schätzungen der Kommunalitäten liefern. Varianzanteile, die eine Variable nur mit einem Faktor gemeinsam hat, können durch die SMCs nicht erfaßt werden. Wir müssen für mittlere Kommunalitäten in der Population bzw. in der Stichprobe um .50 einen maximalen (negativen) Bias erwarten. Für zunehmende Variablenanzahlen kann die Hypothese erstellt werden, daß der Bias geringer wird. Dies ist aber unabhängig von dem Versuchspersonensampling formuliert. Mehr Variablen bei konstanter Versuchspersonenanzahl muß zusätzliche Bias-Effekte haben. Das Verhältnis von Variablenanzahl und Versuchspersonenanzahl hat sich geändert.

3. Schlußfolgerungen aus den bisherigen Betrachtungen

Es gibt in Abhängigkeit von den vier 'main effects', der

Variablenanzahl

Versuchspersonenanzahl

mittleren Kommunalität

Faktorenanzahl

- A. Einen Bereich $\bar{h}^2$, der als Grenzbereich zu bezeichnen wäre. Werte unter dieser Grenze bedeuten, daß in dem faktorenanalytischen Ergebnis Fehlerfaktoren vor den eigentlichen Faktoren in der Eigenwertreihe auftreten.
- B. Einen Bias, einen systematischen Fehler bei der Schätzung der Kommunalitäten der Population durch die Kommunalitäten einer Faktorenanalyse
- C. Einen Bias bei der Schätzung der Kommunalitäten der Population durch die SMCs. Der 'main effect' Faktorenanzahl ist dabei von geringerer Bedeutung
- D. Eine Koppelung systematischer Fehler bei Faktorenanalysen, die SMCs als Kommunalitätenschätzungen verwenden. Da die Auswirkungen dieser Bias entgegengesetzte Vorzeichen haben, werden diese Faktorenanalysen die genauesten Schätzungen liefern.
- E. Eine Koppelung von Bias- Auswirkungen bei der Image - Analyse, wobei die Vorzeichen gleich (negativ) sind:
 - a. SMCs als Kommunalitätenschätzung
 - b. Kovarianzen in den nichtdiagonalen Bereichen der Image - Kovarianzmatrix, die über die multiple Regression gewonnen wurden
- F. Einen Bias auf der Gesamtstichprobenebene im Hinblick auf die mittlere Kommunalität $\bar{h}_A^2$ aus der Analyse wie auch auf der Ebene der einzelnen Variablen und deren h_A^2 .

4. Folgerungen für die Auswertung

4.1. Gesamtstichprobenebene

In Anlehnung an den zu Beginn dargestellten Überblick werden für jede Verfahrensebene, dazu auf der 3.Ebene für jedes faktorenanalytische Verfahren, die mittleren h^2 -Werte erfaßt. Soweit vorhanden, werden mittlere Kommunalitäten auch aus fünf Stichproben gebildet ⁺⁾ .

Stichprobenübersicht

Mittlere Populat. Fakt.lad.	N = 200			N = 100		N = 50	N = 25	n k n/k
	10	10	6	10	15	10	10	
	2	2	2	2	3	2	2	
	5	7,3	3	5	5	5	5	
.20	5	5	5	5			5	
.30	5	5	5	5			5	
.40	5	5	5	5			5	
.50	5	5	5	5	1		5	
.60	1			1	1			
.70	5	5	5	5	1	2	5	
.80	1			1	1			
.90	1			1	1			

Nach dieser Übersicht werden 133 verschiedene Stichproben erfaßt.

Für jede Stichprobe stehen acht verschiedene Kommunalitätenangaben zur Verfügung : Population (h_p^2), Stichprobe (h_S^2), Quadrierter Multipler Korrelationskoeffizient (SMC), Hauptkomponentenanalyse (h_1^2)

⁺⁾ Überla (1968) hat seine Validitätskoeff. aus jeweils 50 Stichproben gewonnen. Eine Überprüfung (Schiller u. Wittmann 1971) hat ergeben, daß sich die Streuungen der Validitätskoeff. und der mittleren Kommunalitäten über einen Stichprobenumfang von fünf Stichproben nicht mehr bedeutsam ändern, das jeweilige Mittel aus 5 Stichproben dem aus 50 Stichproben entspricht.

bzw. (h_{FA1}^2), Hauptachsenanalyse mit Kommunalitätenschätzung durch die SMCs (h_{TFAK}^2), Image-Analyse (h_{IFA}^2), Kanonische Analyse (h_{KFA}^2) und Alpha-Analyse (h_{AFA}^2). Die jeweiligen h^2 -Werte werden je Stichprobe und Analysenverfahren in einen mittleren $\bar{h}^2$ -Wert gefaßt und mit dem Stichproben- h^2 verglichen, um den Betrag eines Bias zu ermitteln. Wo fünf Stichproben vorliegen, kann ein mittlerer Bias bestimmt werden.

Für alle Stichproben werden die Eigenwertreihen aus der TFA1 erfaßt. Über die Validitätskoeffizienten nach der Rotation wird überprüft, in welchen Faktorladungsbereichen Fehlerfaktoren in der Rangreihe der Eigenwerte vor den eigentlichen Faktoren auftreten. Für Stichproben unter dieser Grenze erübrigt sich eine weitere Analyse der Daten.

Ergebnisse aus den Rotationen sind dem vorgesehenen Weg der Datenauswertung nicht zu unterwerfen. Die Kommunalitäten der einzelnen Variablen wie das mittlere $\bar{h}_A^2$ ändern sich nicht.

4.2. Ebene der einzelnen Variablen

Die Idee, Überlaas Validitätskoeffizienten als Korrelation der ursprünglichen mit den geschätzten Faktorwerten nach der Analyse mit den Korrelationen der entsprechenden Faktorladungen zu vergleichen, läßt sich leider nicht verwirklichen. Die Faktorladungen sind, vor allem bei den hier eingehaltenen 'simple-structure' Kriterien, verschiedenen Grundgesamtheiten entnommen. Außerdem wäre auch nur das nach der Rotation erhaltene Faktorladungsmuster einsetzbar. Damit könnten solche Koeffizienten, wie bei Überla, nur eine Aussage über das gesamte Verfahren einschließlich Rotation zulassen.

Dagegen wäre es für alle Ebenen möglich, die jeweiligen h^2 -Werte zu korrelieren. Da in der vorliegenden Untersuchung mit höchstens 15 Variablen zu rechnen ist, kann auch dieser Analysenweg kaum beschrritten werden. Wenn dieser Weg trotzdem eingeschlagen wurde, so geschah dies durchaus in der Kenntnis der Einschränkungen mit dem Ziel, erste Informationen zu der Reproduktion der h^2 zu gewinnen.

Es stellt sich nun die Aufgabe, innerhalb einer Stichprobe die h^2 -Werte der einzelnen Variablen von Ebene zu Ebene zu verfolgen. Der Gesamtstichprobenbias muß dabei berücksichtigt werden. Ein Vergleich von Ebenen im Hinblick auf die einzelnen Variablen läßt sich nur durchführen, wenn die mittleren h^2 jeweils gleich sind. Um dies zu erreichen, wurde folgende Technik eingeschlagen:

Var.NR.	1	2	3		n	$h^2/n = \bar{h}^2$
Stichpr.	h_{1S}^2	h_{2S}^2	h_{3S}^2		h_{nS}^2	$\bar{h}_S^2$
Analyse	h_{1A}^2	h_{2A}^2	h_{3A}^2		h_{nA}^2	$\bar{h}_A^2$
Stichpr.	x_{1S}	x_{2S}	x_{3S}		x_{nS}	: $x_{1S} = h_{1S}^2 - \bar{h}_S^2$
Analyse	x_{1A}	x_{2A}	x_{3A}		x_{nA}	: $x_{1A} = h_{1A}^2 - \text{Bias}_A - h_{1S}^2$

Die x_S geben an, um welchen Betrag die h_S^2 der Variablen vom mittleren $\bar{h}_S^2$ abweichen. Wie schon weiter oben begründet, werden die Analysenergebnisse in Beziehung zu den Stichproben- h^2 gesetzt. Es hat sich gezeigt, daß das $\bar{h}_S^2$ dem $\bar{h}_p^2$ der Population entspricht. Die h_S^2 der einzelnen Variablen zeigen dagegen Abweichungen von den Populationswerten. Da es das Ziel ist, die Fähigkeit einer Analysetechnik zur Reproduktion der eingegebenen Daten zu überprüfen, ist es sinnvoller, sich auf die Stichprobe zu beziehen. Die x_A geben an, ob ein h_S^2 in seiner Abweichung von $\bar{h}_S^2$ entsprechend reproduziert wurde. Dies würde einen $x_A = 0$ ergeben. Positive Werte von x_A bedeuten, daß h_A^2 im Vergleich zu h_S^2 einen erhöhten Wert zeigt, negative x_A bedeuten entsprechend im Vergleich zu niedrige Reproduktion der Kommunalität einer Variable in der Stichprobe. Dabei ist jeweils der Gesamt-Bias berücksichtigt, alle h_A^2 wurden

um den $\text{Bias}_A = \bar{h}_A^2 - \bar{h}_S^2$ vermindert.

Bei der Beurteilung der Ergebnisse kann von der Erwartung ausgegangen werden, daß x_A einen Wert von Null erreichen sollte.

Sind die Vorzeichen von x_S und x_A gleichsinnig, ist die Abweichung der Kommunalität einer Variable vom Mittel nach der Analyse noch größer. Entgegengesetzte Vorzeichen zeigen an, daß eine Abweichung vom Mittel nicht reproduziert wird. Bei gleich hohen Beträgen ist die Variablenkommunalität 'nivelliert' worden. Falls $|x_S| < |x_A|$ bei entgegengesetzten Vorzeichen, ist die in der Stichprobe im Vergleich zu den restlichen Variablen als Variable mit hohem Varianzbeitrag als solche mit eher niedrigem Varianzbeitrag durch die Faktorenanalyse dargestellt.

IV. DIE ERGEBNISSE

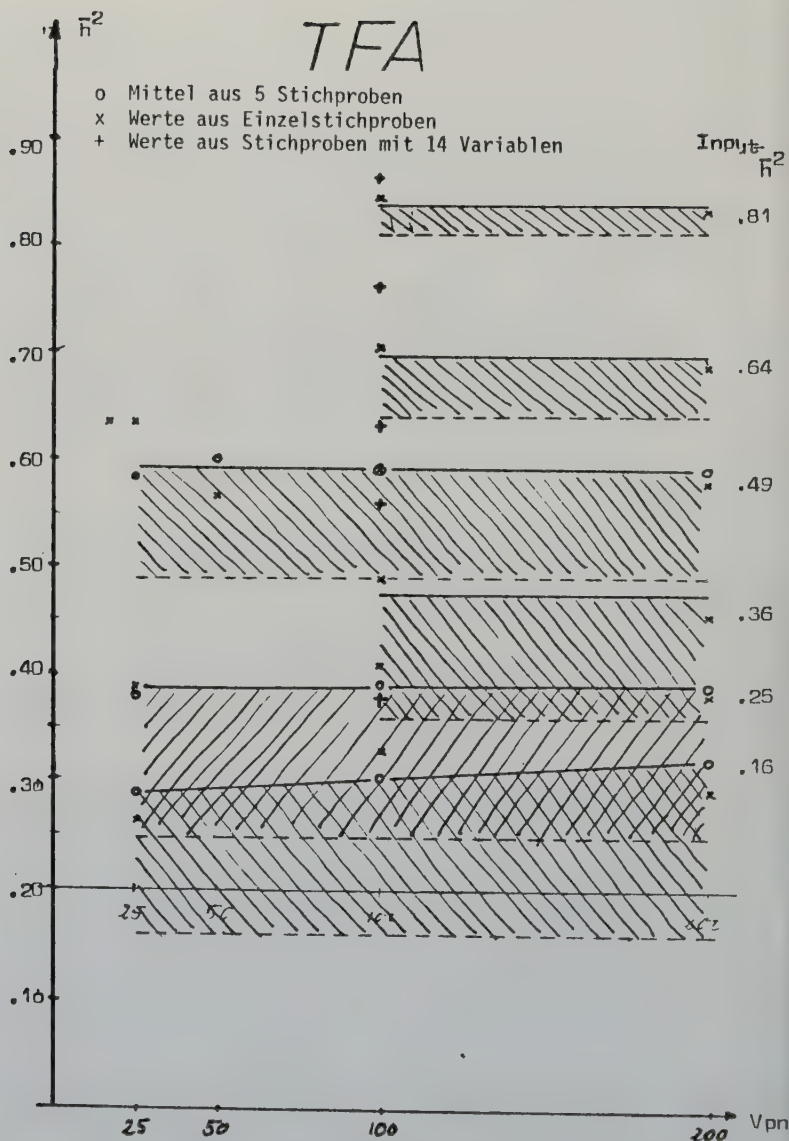
Vorbemerkung: Die in der Folge diskutierten Darstellungen sind aus den im Anhang vorgestellten Daten erarbeitet. Die Daten sind (s. Stichprobenübersicht S.29) Mittelwerte aus 5 Stichproben, soweit dieser Stichprobenumfang vorlag. In manchen Fällen erwies es sich als günstig, Daten für die Darstellungen und Berechnungen mit einer vergleichbaren Basis zu haben, d.h. gleicher mittlerer Stichprobenkommunalität. Dabei wurden dann alle Analysenergebnisse entsprechend dem Umrechnungsverhältnis bei der mittleren Stichprobenkommunalität neu bestimmt (Daten 2). Die Differenzen aus den mittleren Kommunalitäten der Analysen und der Stichprobe sind entsprechend aus den 'Daten 2' erarbeitet (Daten 3).

1. Der systematische Fehler auf der Stichprobenebene

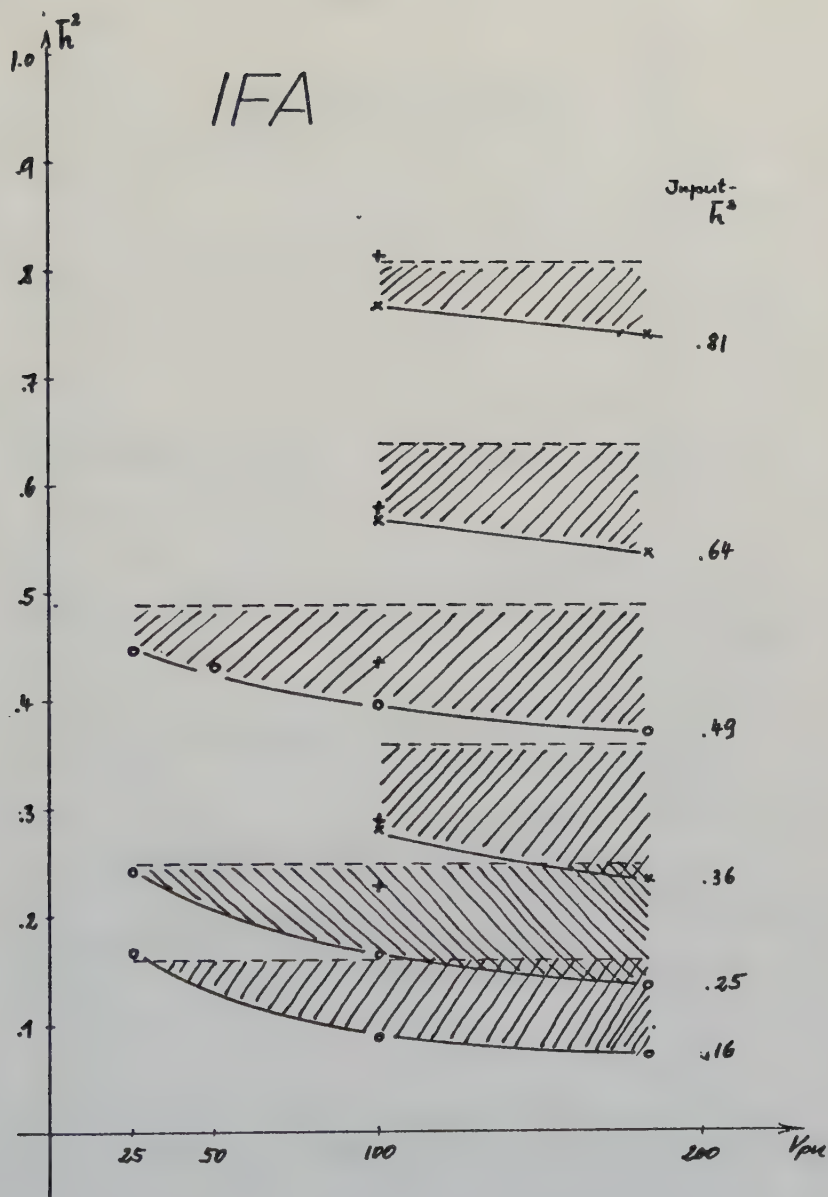
1.1. Die Bedeutung der Versuchspersonenanzahl

1.1.1. Überblick über die Ergebnisse bei den verschiedenen Analysenverfahren

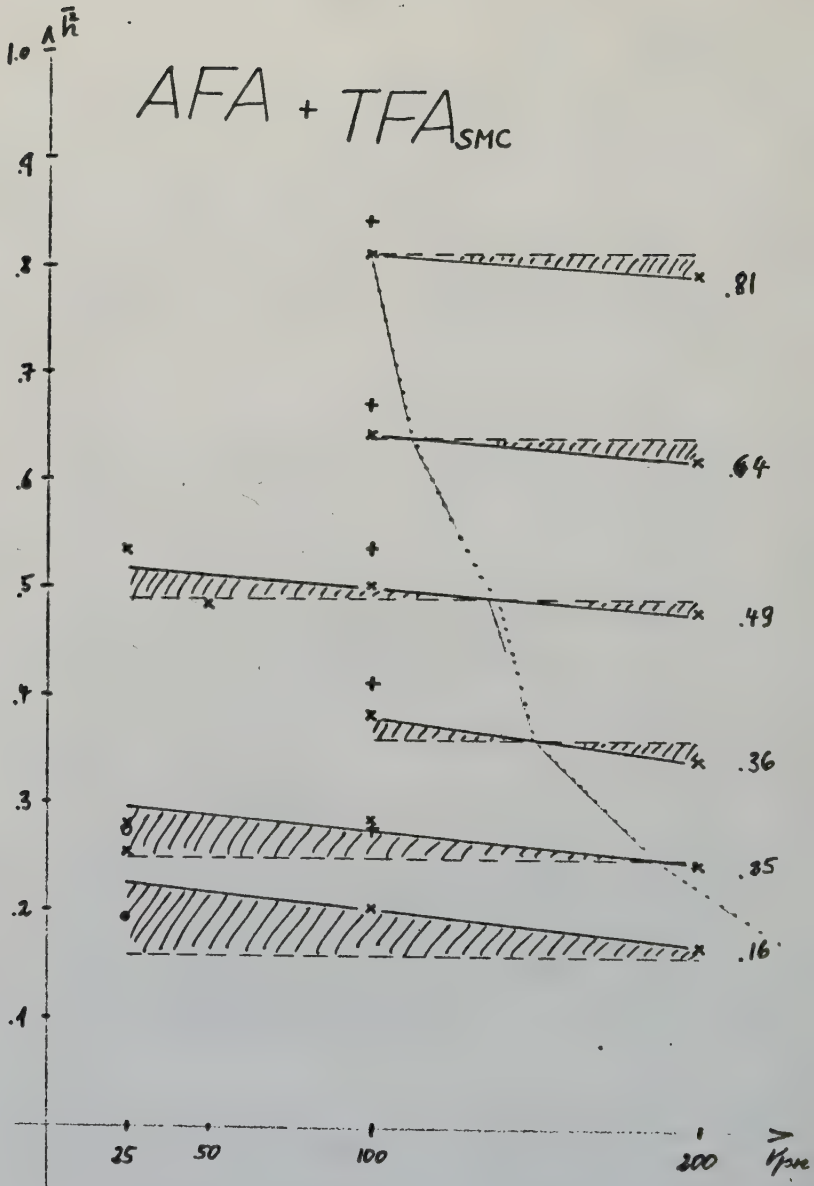
Diese Ergebnisse sind in den Darstellungen 2 bis 6 auf den folgenden Seiten zu finden. Für diese Darstellungen gilt: Die Output-Varianz $\bar{h}_A^2$ wird in Abhängigkeit von dem Stichprobenumfang N dargestellt. In der Hauptsache handelt es sich um die Kurven der Stichproben mit zehn Variablen. Die einzelnen Kurven beziehen sich jeweils auf gleiche Input- Ladungen bzw. gleiche mittlere Input-Kommunalitäten $\bar{h}_S^2$ in der Stichprobe. Gestrichelt ist jeweils die Gerade $\bar{h}_{\text{Input}}^2 = \bar{h}_{\text{Output}}^2$ eingezeichnet. Sie beschreibt die erwarteten Ergebnisse. Die schraffierten Flächen kennzeichnen den Umfang des systematischen Fehlers, und zwar bei der Faktorenanalyse mit der Vorkenntnis der 'wahren' Anzahl der Faktoren. Ohne diese Bedingung entstanden die SMC- Ergebnisse.



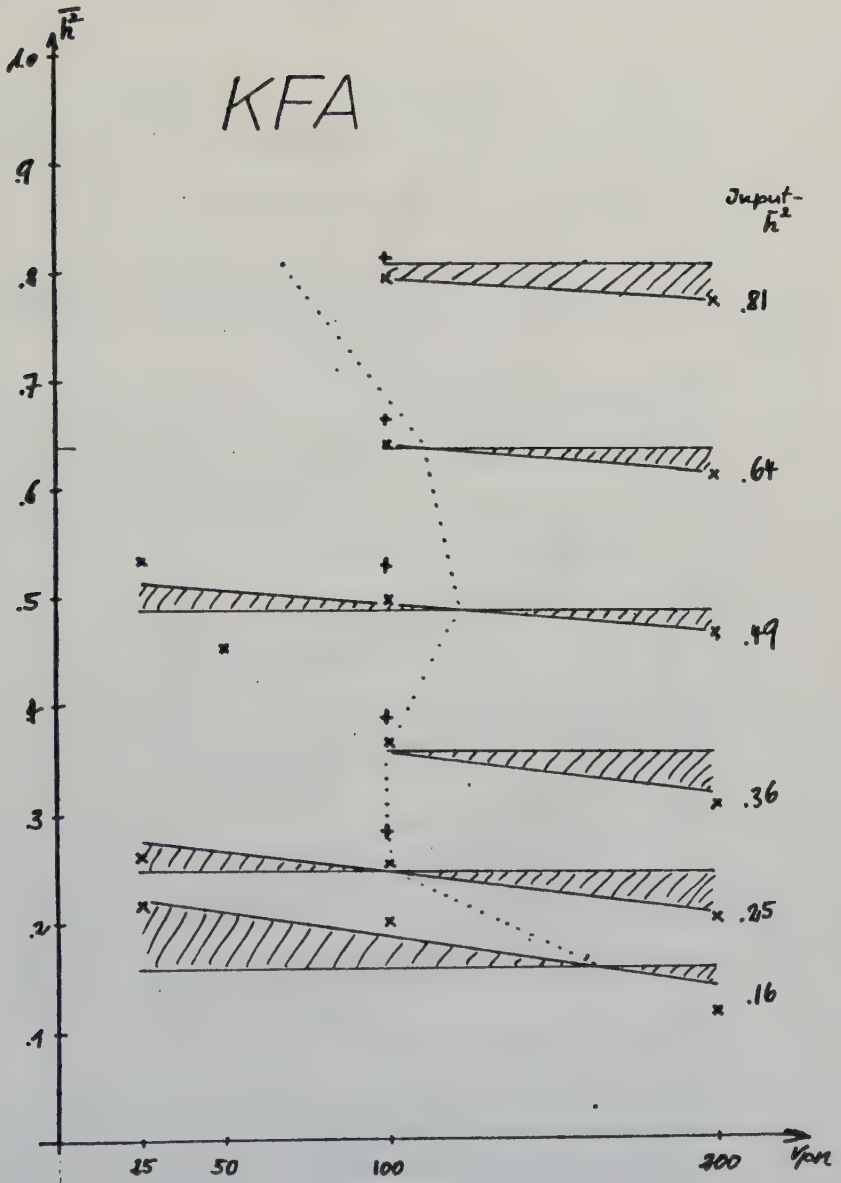
Darstellung 2 : Mittlere Kommunalitäten der Analyseergebnisse der TFA1 in Abhängigkeit von der Vpn-Anzahl N



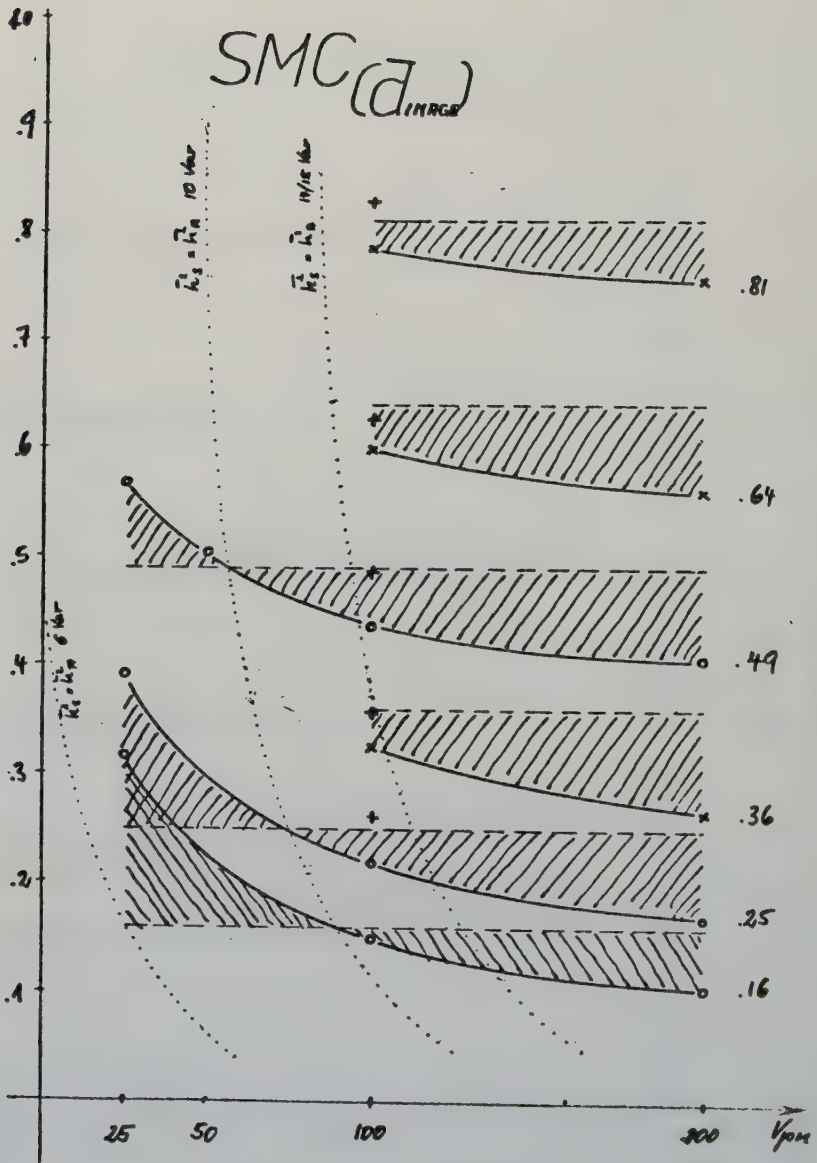
Darstellung 3 : Mittlere Kommunalitäten aus der Image-Analyse
in Abhängigkeit von der Vpn-Anzahl N



Darstellung 4 : Mittlere Kommunalitäten aus der AFA bzw. TFAK
in Abhängigkeit von der Vpn-Anzahl N



Darstellung 5 : Mittlere Kommunalitäten aus der KFA
in Abhängigkeit von der Vpn-Anzahl N



Darstellung 6 : SMC als mittlere Kommunalitätsschätzung
in Abhängigkeit von der V_{pn} -Anzahl

A. Hauptkomponentenanalyse TFA1 (Darstellung 2 S.34)

Die TFA1-Ergebnisse kennzeichnend und von allen anderen Analyseverfahren abhebend ist die Tatsache, daß bei festen mittleren Kommunalitäten im Input der systematische Fehler bei verschiedenen Versuchspersonenanzahlen fast gleich bleibt. Abweichungen sind nur im Bereich niedriger Varianzbeiträge der Variablen zum Faktorenum (16% und niedriger) feststellbar. Hier zeigt sich ein eher geringerer Bias. Wie später noch zu zeigen sein wird, handelt es sich hier um den Grenzbereich, in dem Fehlerfaktoren dieses Ergebnis bestimmen.

Der Bias ist über die verschiedenen $\bar{h}^2$ des Input bzw. der Stichprobe positiv, mit abnehmender Tendenz von den niedrigen zu den hohen $\bar{h}_S^2$ -Werten.

B. Image - Analyse IFA (Darstellung 3 S.35)

Der Bias ist größtenteils negativ, bis auf einen Bereich niedriger Versuchspersonenanzahl und gleichzeitig geringer mittlerer Input-Kommunalitäten.

Für zunehmende Vpn-Anzahlen bei festem $\bar{h}_S^2$ strebt der Bias gegen einen Grenzwert.

Der Betrag des Bias ist im mittleren $\bar{h}_S^2$ -Bereich am größten.

C. Hauptachsenanalyse TFAK bzw. TFA_{SMC}

Alpha - Faktorenanalyse AFA

Kanonische Analyse KFA (Darstellungen 4 u. 5 S.36/37)

Die Ergebnisse stammen im Gegensatz zu TFA1 und IFA aus Einzelstichproben. Es gibt aber keinen Grund, für mittlere Werte aus 5 Stichproben große Differenzen zu den dargestellten Ergebnissen zu erwarten. Dabei ist auch daran zu denken, daß ja jeweils mittlere Kommunalitäten erarbeitet wurden.

Für alle drei Analysetechniken gilt, daß der Bias im Vergleich zu IFA und TFA1 relativ gering ist. Bei geringer Vpn-Anzahl zeigen sich positive, bei hoher Vpn-Anzahl negative Bias-Werte.

Zunehmende $\bar{h}_S^2$ -Werte bei fester Vpn-Anzahl bewirken niedrigere positive bzw. höhere negative Bias-Werte.

Für alle drei Verfahren zeigt der Schnitt mit der $\bar{h}_S^2$ -Ebene, d.h. wenn der Bias = 0 ist, etwa den Verlauf eines Hyperbelastes einer rechtwinkligen Hyperbel. (s. gepunktete Linie in den Darstellungen)

Es läßt sich in Anlehnung an die Ergebnisse der IFA vermuten, daß der Bias für zunehmende Vpn-Anzahlen einem negativen Grenzwert zustrebt.

D. Quadrierte multiple Korrelationen $\overline{SMC}$ (Darstellung 6 S.38)

Der Bias nimmt bei niedriger Vpn-Anzahl positive Werte ein. Mit zunehmender Vpn-Anzahl wird der Bias negativ und strebt einem negativen Grenzwert zu.

Im Hinblick auf diesen Grenzwert ist der Bias in seinem Betrag am höchsten bei mittleren Werten von $\bar{h}_S^2$. Zunehmende bzw. abnehmende $\bar{h}_S^2$ -Werte bewirken eine Abnahme des Betrages des Bias. Eine entsprechende Aussage war auch bei der IFA gegeben. Die $\overline{SMCs}$ erweisen sich, im Hinblick auf die Ergebnisse aus den Faktorenanalysen, am empfindlichsten auf eine Änderung der Versuchspersonenanzahl im Bereich niedriger Vpn-Anzahlen.

Die in der Darstellung gepunkteten Kurven stellen den Bereich dar, in dem jeweils für feste Variablenanzahlen $\overline{SMC} = \bar{h}_S^2$. Mit anderen Worten: Die Kurven zeigen, bei welcher Vpn-Anzahl der $\overline{SMC}$ so groß ausfällt wie $\bar{h}_S^2$, jeweils bei fester Variablenanzahl. Aus den Ergebnissen der Stichproben mit 6 bzw. 14/15 Variablen sind jeweils die Näherungskurven zusätzlich dargestellt.

An dieser Stelle der Analyse der vorliegenden Daten schien das Ziel, das dieser Arbeit gesetzt war, erreicht zu sein. Es sollten Funktionen angegeben werden, die das Ausmaß der Anpassung an die 'wahre' Struktur (hier: mittlere Kommunalität) angeben. Außerdem sollte das 'blinde' Erfassen der Anzahl der Faktoren dabei ermöglicht werden. Beide Bedingungen oder Forderungen erfüllten die $\overline{SMC}$ -Kurven: Wenn sich diese Kurven in Abhängigkeit von den 'main effects', der Versuchspersonenanzahl, der Variablenanzahl und der mittleren Input - Kommunalität, ohne Abhängigkeit von der

Faktorenanzahl, als Gleichungen beschreiben lassen, dann kann daraus das unbekannte $\bar{h}_S^2$ berechnet werden. Damit hätte man

- eine Schätzung der mittleren Stichprobenkommunalität
- damit auch eine Schätzung der Varianz, die die Faktoren ausschöpfen müssen
- damit eine Bestimmung der Faktorenanzahl, die einen Bias in Abhängigkeit von den 'main effects' ausklammert

Bei genauerer Betrachtung war erkennbar, daß es sich bei den Kurven der Darstellung 6 um Hyperbeläste vermutlich rechtwinkliger Hyperbeln handelt. Für den Praktiker, der mit der Faktorenanalyse empirisch gewonnene Daten beurteilen muß, wären nur die Formeln für die einzelnen Hyperbeln zu bestimmen:

- Hyperbelgleichungen für die $\bar{S}MC$ s in Abhängigkeit von $\bar{h}_S^2$, N und n
- Hyperbelgleichungen für die Mindestgröße der $\bar{S}MC$ s, um noch sichere Lösungen zu bekommen, ebenfalls in Abhängigkeit von $\bar{h}_S^2$, N und n. Dabei wird erwartet, daß es sich bei diesen Gleichungen um die Beschreibung von Hyperbelästen ähnlich den $\bar{h}_S^2 = \bar{S}MC$ (= $\bar{h}_A^2$) handelt.

Diese Hyperbelgleichungen wurden erarbeitet für n = 10. Es würde zu weit führen, diese hier darzustellen und zu diskutieren. Eine Erwartung, die bei der Erarbeitung der Gleichungen bestand, wurde außerdem nicht erfüllt. Die Hoffnung war, daß bei einem Vergleich der Hyperbeln für jeweils feste $\bar{h}_S^2$ sich jeweils gleiche Strukturen (Krümmung im Scheitel) zeigen würden, nur die Asymptoten für $N \rightarrow \infty$ sich je nach $\bar{h}_S^2$ ändern würden. Die Möglichkeit für den Praktiker ein Spektrum von Gleichungen für jeweils variierende $\bar{h}_S^2$ und n zu liefern, wurde verworfen. Es wären auch bei einer Vielzahl von solchen Hyperbelgleichungen jeweils nur diskrete Punkte bzw. Punktescharen beschreibbar gewesen. Ähnliches gilt für eine tabellarische Erfassung. Eine Zusammenfassung der Gleichungen für jeweils festes n war ebenfalls nicht möglich. Darstellung 6 zeigt, daß damit die stark schwankenden Asymptoten nicht berücksichtigt worden wären, eine Schätzung von $\bar{h}_S^2$ hier gleich wieder mit einem Fehler belegt worden wäre.

Ein erster Hinweis, daß es noch eine elegantere Lösung des Problems geben müsse, waren die Hyperbeln für $\overline{SMC} = \overline{h}_S^2$ (Darstellung 6 S.38, gepunktete Kurven). Hier deutete sich eine formale Abhängigkeit des 'Variablensampling' und des 'Versuchspersonensampling' ab. Mit anderen Worten: Die Abstände der Hyperbeläste für $\overline{SMC} = \overline{h}_S^2$ bei verschiedenen Variablenanzahlen müssen dazu dienen können, die $\overline{h}_S^2$ - Schätzungen aus dem Bereich für 10 Variablen in Bereiche beliebiger Variablenanzahlen umsetzen zu können.

Bevor dieser Ansatz ausgenutzt wird, sollen noch eine Reihe weiterer Analysen der Ergebnisse erfolgen.

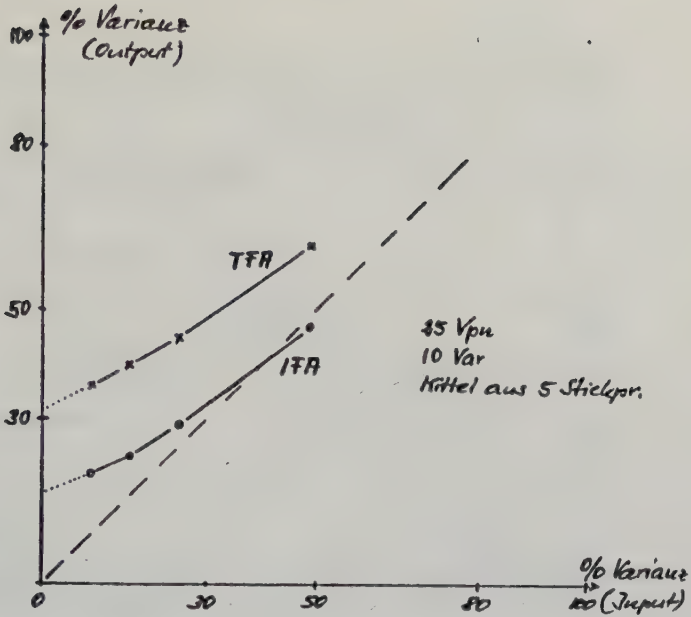
1.2. Die Bedeutung der mittleren gemeinsamen Varianz in der Stichprobe

Es muß an dieser Stelle noch einmal darauf hingewiesen werden, daß die Stichproben erwartungstreu Schätzungen der $\overline{h}_S^2$ liefern. Diese Tatsache ermöglicht es, in der Folge sich auf die Vergleiche der FA-Ergebnisse mit den Stichproben zu beschränken. Diese Beschränkung muß allerdings bei der Betrachtung der Ergebnisse auf der Ebene der einzelnen Variablen neu überdacht werden.

Es hat sich bisher gezeigt, daß TFA1 und IFA (s. Darstellung 2 bis 6) Extrempositionen im Hinblick auf den Bias, den systematischen Fehler, einnehmen. Die TFA1 liefert die höchsten und ausschließlich positiven, die IFA die höchsten negativen systematischen Fehler. AFA, KFA und TFAK bewegen sich dazwischen und sind in ihren Ergebnissen kaum unterscheidbar. Die $\overline{SMC}$ - Verläufe (s. auch Darstellung 1) können grob als zwischen diesen drei Verfahren und der IFA sich bewegend beschrieben werden.

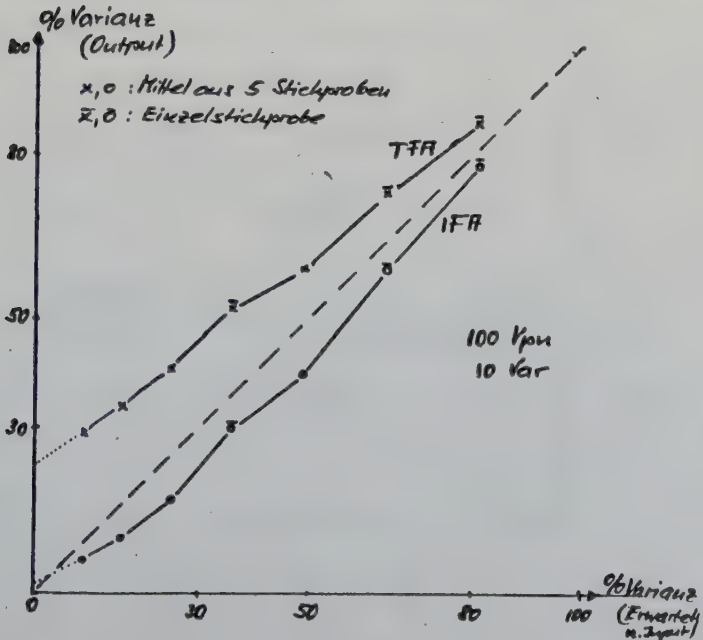
1.2.1. Varianzausschöpfung in den einzelnen Analysentechniken

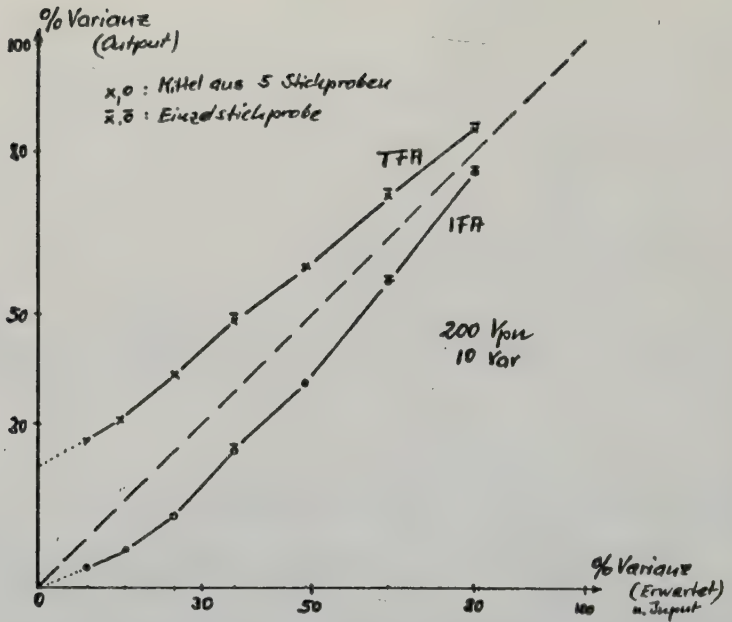
Die Darstellungen 7 bis 9 auf den folgenden Seiten zeigen den systematischen Fehler in Abhängigkeit von der Input-Varianz $\overline{h}_S^2$. Die Versuchspersonen- und Variablenanzahlen sind jeweils fest. Die gestrichelte Linie beschreibt die erwarteten Ergebnisse als nicht bias-behaftete Ergebnisse.



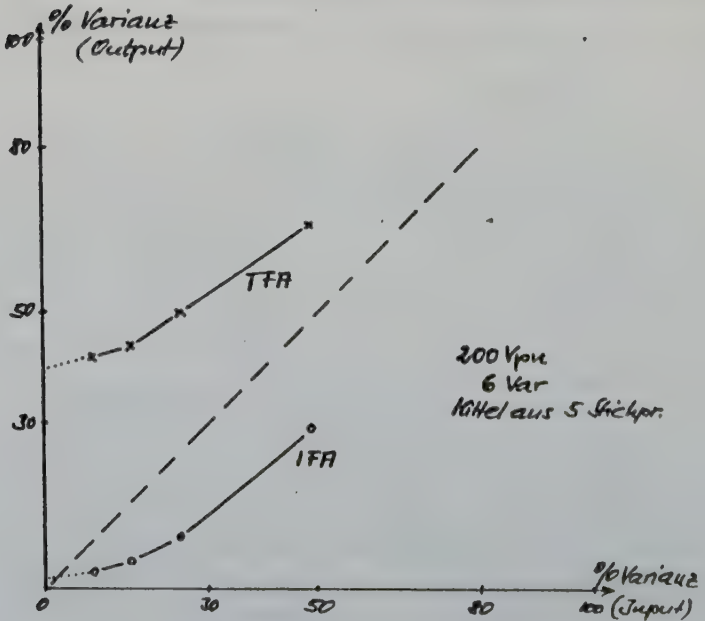
Darstellung 7 : Varianzausschöpfung durch TFA1 und IFA

Darstellung 8 : Varianzausschöpfung durch TFA1 und IFA





Darstellungen 9 u. 10 : Varianzeusschöpfung durch TFA₁ und IFA



Als Kennzeichnung der Ergebnisse kann festgehalten werden:

Für die TFA1:

- A. Die TFA1 liefert einen fast linear von der mittleren Uniqueness abhängigen Bias.

$$\text{Bias}_{\text{TFA1}} = f (1 - \bar{h}_S^2)$$

- B. Der TFA1 - Bias wird geringer mit zunehmender Vpn-Anzahl

$$\text{Bias}_{\text{TFA1}} = f (N ; 1 - \bar{h}_S^2)$$

- C. Der Bias wird größer bei geringerer durchschnittlicher Variablen-anzahl je Faktor

$$\text{Bias}_{\text{TFA1}} = f (N ; n/k ; 1 - \bar{h}_S^2)$$

Für die IFA:

- A. Der IFA - Bias zeigt eine eher kurvilenare Abhängigkeit von der mittleren Kommunalität der Stichproben

$$\text{Bias}_{\text{IFA}} = f (\bar{h}_S^2)$$

- B. Für mittlere $\bar{h}_S^2$ ergeben sich die höchsten (negativen) Bias-Werte

$$\text{Bias}_{\text{IFA}} = f (| \bar{h}_S^2 - .50 |)$$

- C. Der Betrag des Bias nimmt für steigendes N zu

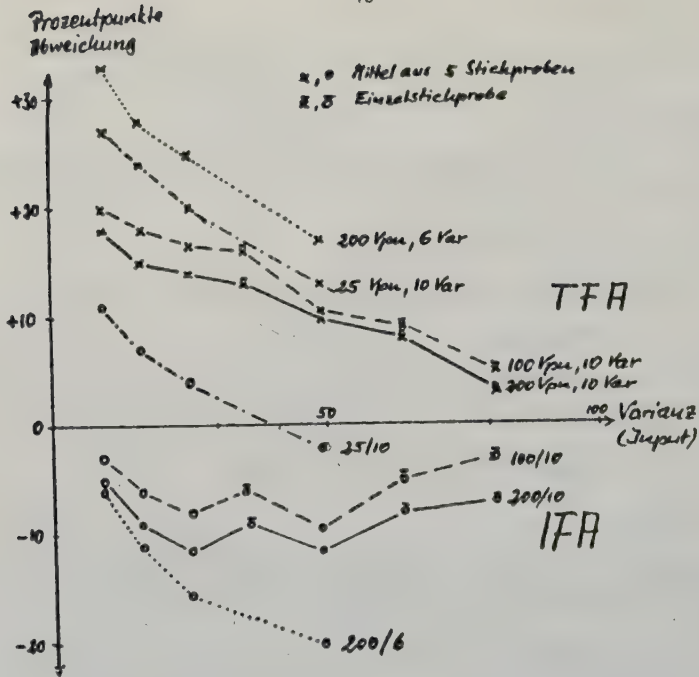
$$\text{Bias}_{\text{IFA}} = f (N ; | \bar{h}_S^2 - .50 |)$$

- D. Bei gleicher Vpn-Anzahl steigt der Bias bei abnehmender Anzahl der Variablen je Faktor

$$\text{Bias}_{\text{IFA}} = f (N ; n/k ; | \bar{h}_S^2 - .50 |)$$

1.2.2. Der systematische Fehler in Abhängigkeit von der mittleren Input-Kommunalität (TFA1 und IFA)

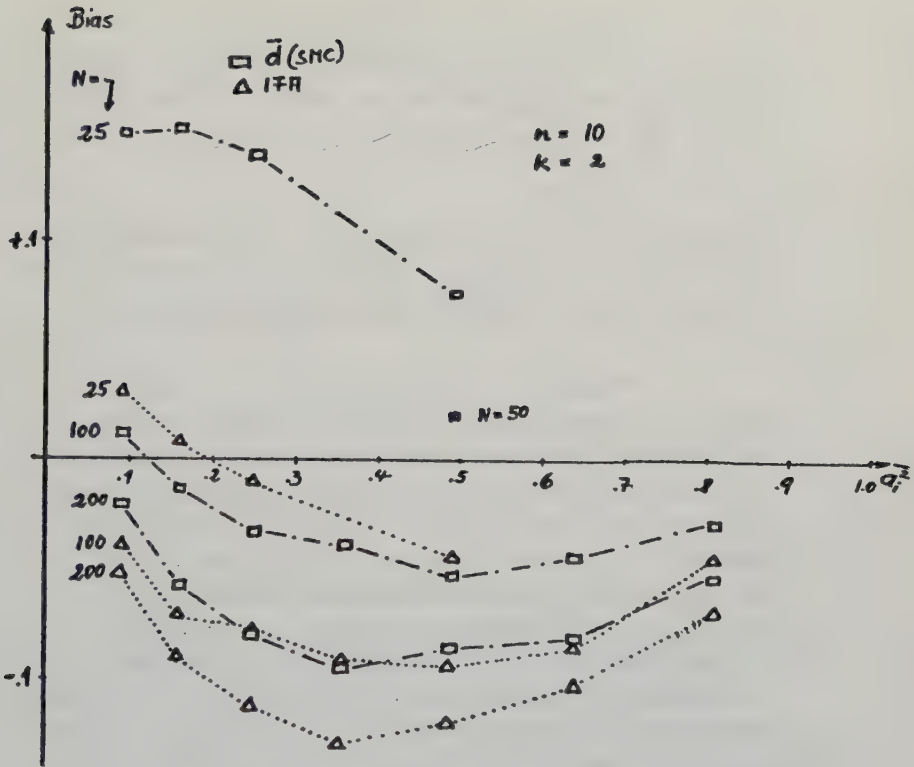
Die Darstellung 11 der nächsten Seite ist noch einmal eine Zusammenfassung der Darstellungen 7 bis 10, allerdings in der Form, daß nur



Darstellung 11: Der Bias bei TFA1 und IFA

die Abweichungen von der erwarteten Ausschöpfung der Input-Varianz, $\bar{h}_A^2 - \bar{h}_S^2$, prozentual dargestellt werden. Es bestätigen sich die Aussagen, die bisher gewonnen wurden, für beide Analysenverfahren. Bemerkenswert sind die Einzelstichprobenergebnisse aus der Image-Analyse. Bei diesen ersten Stichproben aus der Serie scheint es zufällig mehr multiple Kovarianzen gegeben zu haben.

Um ein einheitliches Bild der Bias-Verläufe zu erhalten, soll noch eine Darstellung allein aus den Ergebnissen der ersten Stichprobe folgen. Dabei werden IFA und $\bar{SMC}$ verglichen. (Darstellung 12) Hier wird bewußt in Kauf genommen, daß diese ersten Stichproben ein 'Mehr' an Kovarianz aufweisen, als es über ein Mittel aus fünf Stichproben zu erwarten wäre. Da es aber um die Erarbeitung eines geschlossenen Bildes, um den Verlauf des Bias über die gesamte Breite der möglichen Input-Varianz geht, kann vorläufig darüber hinweggesehen werden. Besonderes Augenmerk soll den SMC-Bias



Darstellung 12: Der Bias bei IFA und $\bar{SMC}$

zukommen, da nach den bisherigen Erkenntnissen über die SMCs eine Lösung der angestrebten Ziele zu erwarten ist.

1.2.3. Der systematische Fehler in Abhängigkeit von der mittleren Input-Kommunalität (IFA und SMC)

Eine Analyse der Darstellung 12 läßt auf den ersten Blick folgende Aussagen zu:

- $\bar{SMC}$ - Bias und $\bar{h}_{IFA}^2$ - Bias zeigen parallele Bewegungen für $N = 100$ und $N = 200$. Hier gelten die Schlußfolgerungen, die schon für den IFA - Bias getroffen wurden.

B. Der IFA - Bias ist in seinem Betrag etwas höher. Dies entspricht den Erwartungen aus der Theorie der Image-Analyse.

C. Der $\overline{SMC}$ - Bias reagiert empfindlicher auf geringe V_{pn} - Anzahlen.

Besonders der $\overline{SMC}$ - Bias bei $N = 25$ gab einige Probleme auf. Die oben unter A, getroffenen Feststellungen ließen vermuten, daß sich der $\overline{SMC}$ - Bias wie bei der IFA darstellen ließe als :

$$\text{Bias}_{SMC} = f (N ; n ; \overline{h}_S^2 - .50)$$

Dabei ist n/k durch n ersetzt. Je mehr Variablen pro Faktor in der Stichprobe, desto höher werden die SMCs sein. Sie berücksichtigen die 'Dichte' der Variablen im Raum der gemeinsamen Varianz und können daher unabhängig von der Faktorenanzahl dargestellt werden.

Der Bias bei $N = 25$ läßt kaum eine Abhängigkeit von $\overline{h}_S^2 - .50$ erkennen. Ein Erklärungsversuch über die hier bedeutsamer werdenden Fehlerfaktoren ist auch nur in geringem Umfang zutreffend. Eine Lösung des Problems brachte eine genauere Analyse der Kurvenverläufe bei $N = 200$ und $N = 100$.

Wenn

$$\text{Bias}_{SMC} = f (\overline{h}_S^2 - .50)$$

stimmt, dann ist nicht erklärbar warum

- die Kurven nicht voll symmetrisch zu einer Achse senkrecht auf der $\overline{h}_S^2 -$ Koordinate durch .50 verlaufen
- die Kurve für $N = 100$ sogar die $\overline{h}_S^2 -$ Achse schneidet.

Die Lösung muß darin liegen, daß der $\overline{SMC}$ - Bias durch zwei verschiedene Bias - Effekte (bei festem N und n) bestimmt ist. Dabei scheint einer dieser Effekte positive Bias - Werte ähnlich derer bei der IFA1 zu liefern. Wenn dieser Einzeleffekt bei geringer Versuchspersonenanzahl groß wird gegenüber dem zweiten Effekt, dann ist daraus auch der Kurvenverlauf für $N = 25$ erklärbar.

Eine graphische Darstellung dieser Schlußfolgerungen zeigt die folgende Abbildung.

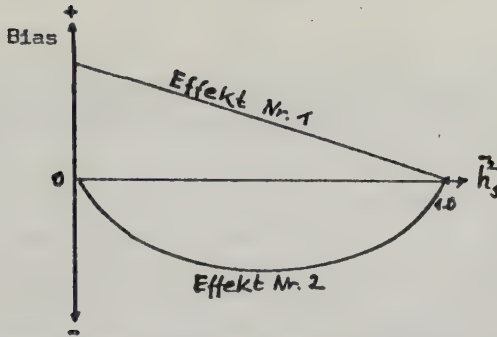


Abbildung 1: Aufspaltung des $\overline{SMC}$ - Bias in Bias - Effekte

Daraus kann eine Bestimmungsgleichung, vorläufig ohne Berücksichtigung von N und n , abgeleitet werden :

$$\text{Bias}_{\text{SMC}} = f \left(1 - \overline{h}_S^2 ; \overline{h}_S^2 - .50 \right)$$

Wie groß jeweils die beiden Bestimmungsgrößen ausfallen, wird dann von der Versuchspersonen- und Variablenanzahl abhängen.

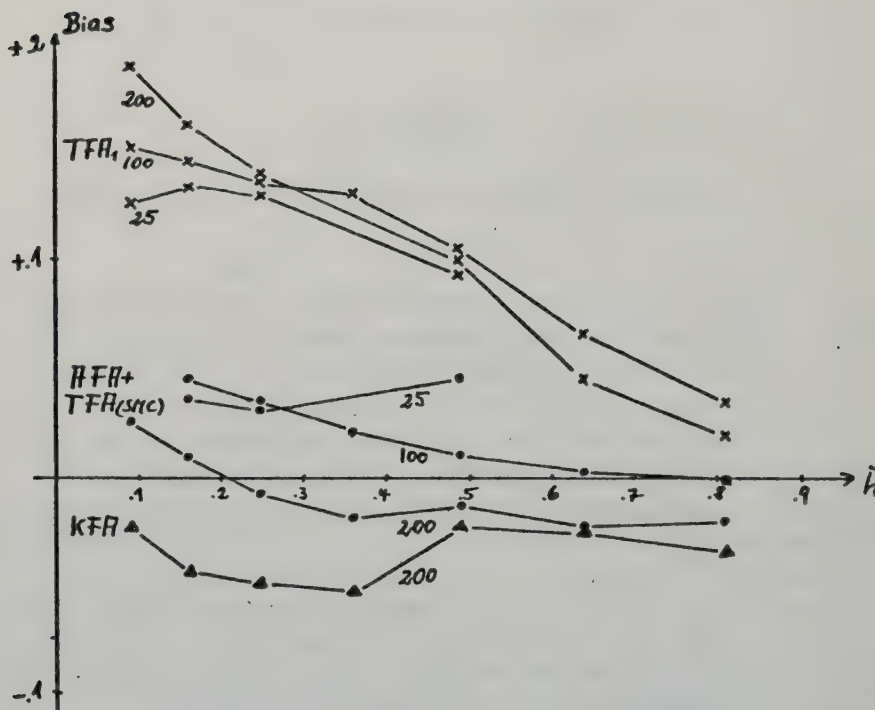
Eine Interpretation der beiden Bias - Effekte im $\overline{SMC}$ - Bias könnte lauten:

Bias - Effekt Nr. 1 : Es handelt sich um einen im Wert positiven Bias. Hier wird zusätzliche gemeinsame Varianz als Folge von Interkorrelationen der gemeinsamen und unigen Anteile der Variablen, wechselseitig und untereinander, ausgeschöpft.

Bias - Effekt Nr. 2 : Es handelt sich um einen im Wert negativen Bias. Die Varianz, die eine Variable nur mit einem Faktor gemeinsam hat, kann durch die SMCs nicht beschrieben werden. Die SMCs werden den Varianzanteil des Faktorenraumes unterschätzen.

Die Interpretationen der beiden Bias - Effekte bei den $\overline{SMC}$ fügen sich sinnvoll in die bisherigen theoretischen Betrachtungen und erklären die aus den Simulationen gewonnenen Ergebnisse. Die obige allgemeine Formel müßte noch in ihrer Abhängigkeit von den weiteren 'main effects', der Variablen- und Versuchspersonenanzahl bestimmt werden.

1.2.4. Der systematische Fehler in Abhängigkeit von der mittleren
Input- Kommunalität (TFA_{SMC} , AFA , KFA)



Darstellung 13: Der Bias bei TFAK , AFA und KFA

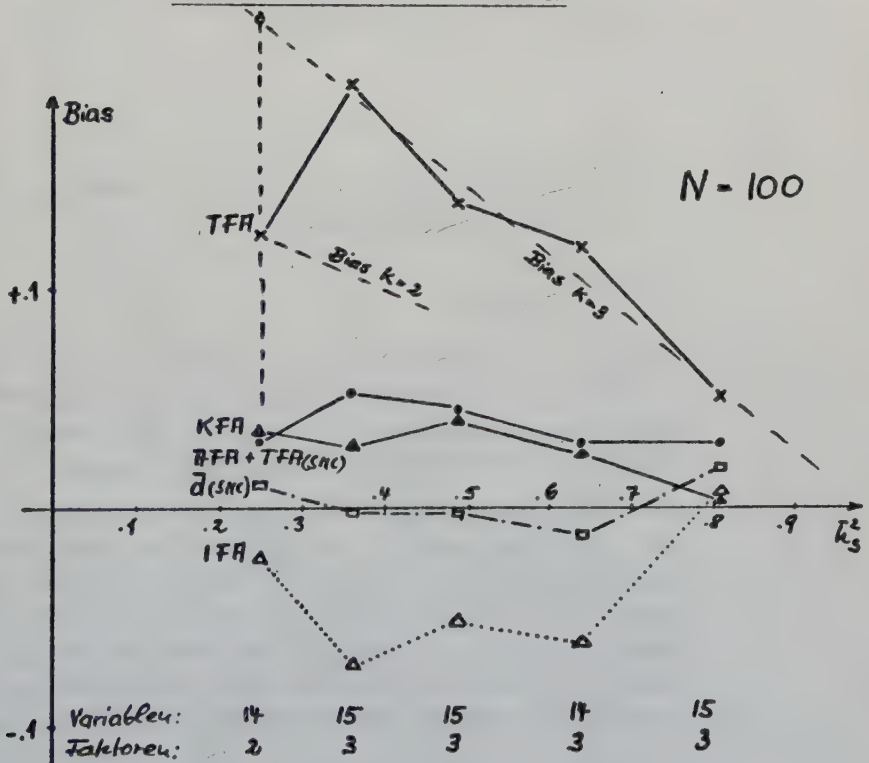
Zum Vergleich sind in dieser Darstellung auch die TFA1 - Ergebnisse aufgeführt. Dabei handelt es sich im Unterschied zu Darstellung 11 um Stichproben- korrigierte Werte. Alle Ergebnisse wurden auf $\bar{h}_S^2$ umgerechnet. Diese geringfügig veränderten Ergebnisse sollen hier nicht zur Debatte stehen.

Die Ergebnisse aus TFA1 , AFA und KFA aus Einzelstichproben sind insgesamt gekennzeichnet durch geringere Varianzabweichungen von den 'wahren' Beträgen. Für den Praktiker kann der Schluß gezogen werden, daß diese faktorenanalytischen Verfahren die ursprünglichen Kommunalitäten am besten reproduzieren, allerdings unter der Voraussetzung, daß die 'wahre' Anzahl der Faktoren bekannt ist.

Die KFA - Ergebnisse sind für $N = 200$ gesondert aufgeführt. Bei $N = 100$ und $N = 25$ unterscheiden sich die drei Analysenverfahren kaum.

Im Vergleich zu der TFAI kann die starke Annäherung an die Stichprobenkommunalitäten begründet werden durch die diesen Verfahren eigene Kommunalitätsschätzung über die SMCs. Die Kurven zeigen auch jeweils deutlich deren Einfluß.

1.2.5. Der systematische Fehler in den verschiedenen Analyse- techniken bei 14 bzw. 15 Variablen



Darstellung 14 : Der Bias bei Stichproben mit 14 / 15 Variablen

Die schon bisher erarbeiteten Ergebnisse bestätigen sich auch hier. IFA und TFA1 beschreiben die jeweiligen Extreme eines möglichen Bias. Die $\overline{SMC}$ - Kurve zeigt eine starke Annäherung an die 'wahren' Werte, TFAK, AFA, und KFA liefern kaum unterscheidbare Ergebnisse.

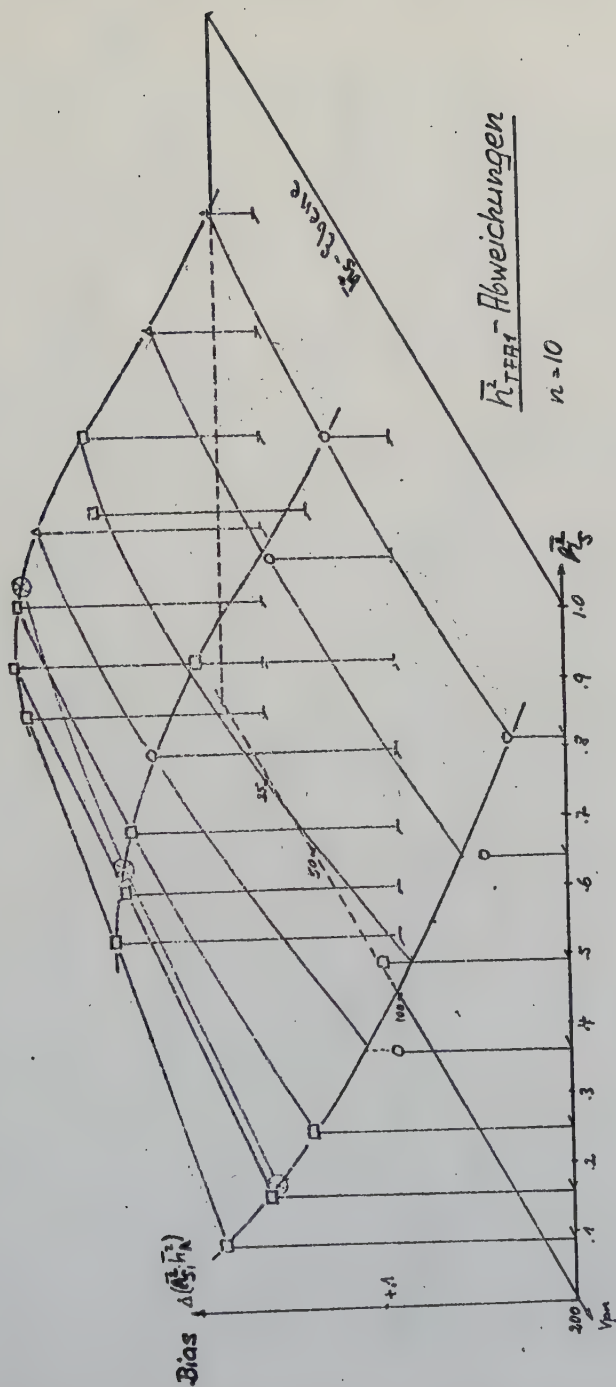
Das bemerkenswerteste Ergebnis ist der Unterschied der TFA1 - Werte der zweifaktoriellen Stichprobe gegenüber den Stichproben mit drei Faktoren. Der Bias der TFA1 hängt wesentlich von der Anzahl der extrahierten Faktoren ab. Die gestrichelten Geraden sollen dies deutlich machen. Je Faktor wird, bei festem N, n und $\overline{h}_S^2$, ein konstanter Bias - Betrag zusätzlich durch $\overline{h}_{TFA1}^2$ geboten. Für jede spätere Bestimmung der Faktorenanzahl aus der Eigenwertreihe der TFA1 muß diese Tatsache berücksichtigt werden.

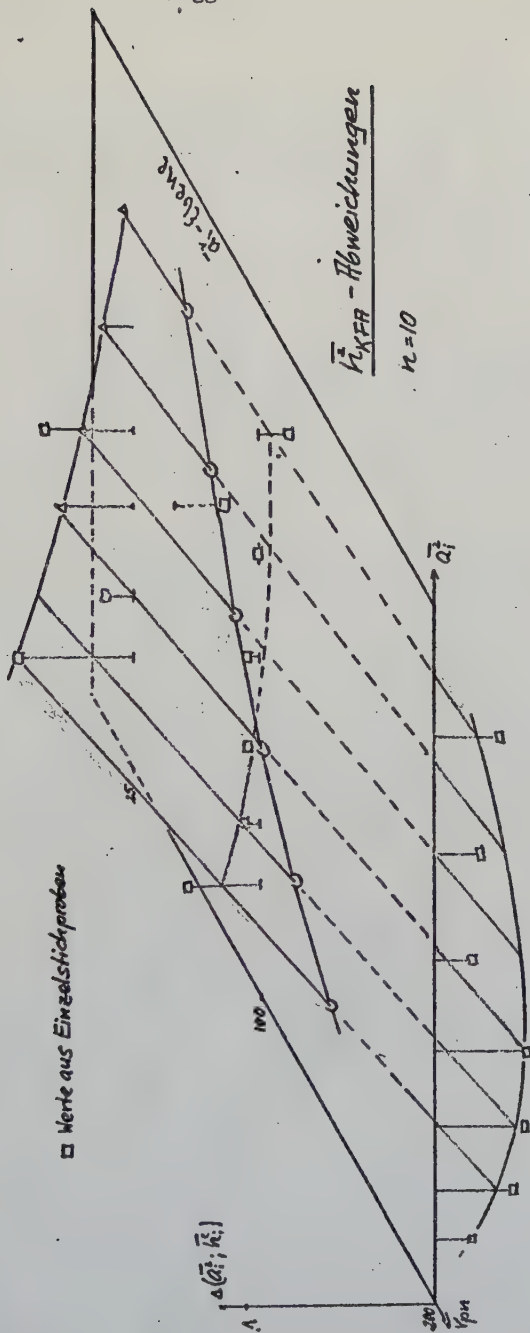
1.3. Der systematische Fehler als Funktion von gemeinsamer Varianz und Versuchspersonenanzahl

1.3.1. Vorbemerkungen

Die jetzt zur Darstellung kommenden Ergebnisse sind jeweils aus den Analysen mit zehn Variablen, dem Schwerpunkt der vorliegenden Simulationen, gewonnen. Für TFA1, IFA und SMC sind jeweils, soweit vorhanden, die Mittelwerte aus 5 Stichproben eingesetzt.

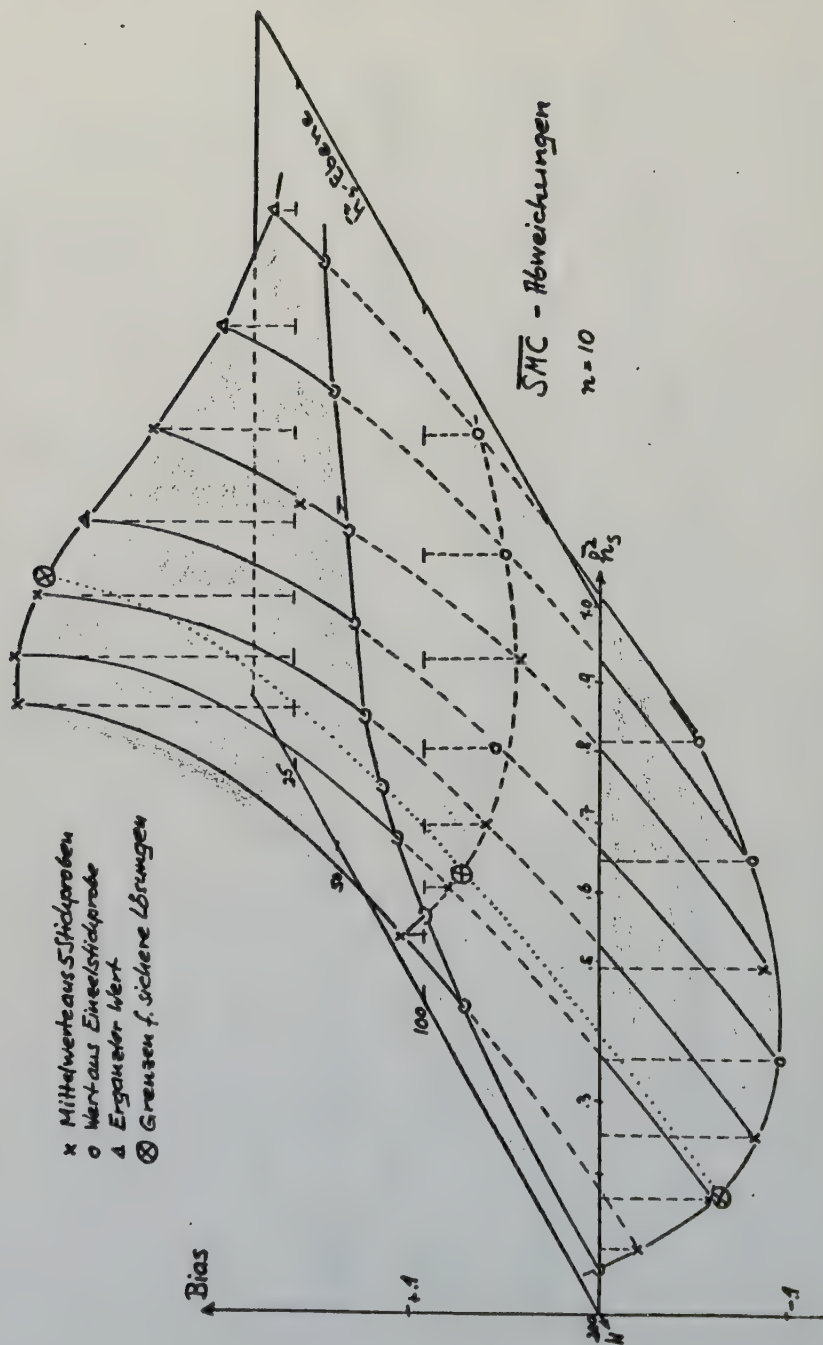
AFA, KFA und TFAK (=TFA_{SMC}) sind beispielhaft nur durch die KFA - Ergebnisse vorgestellt. Die Ergebnisse aus diesen Verfahren (abgesehen davon, daß sie jeweils nur aus einer Stichprobe gewonnen wurden) sind für die weitere Erarbeitung von geringerer Bedeutung. Dadurch soll keineswegs ihre Bedeutung als beste Verfahren zur Schätzung der mittleren Input-Kommunalität, unter der Voraussetzung einer bekannten Faktorenanzahl, geschmälert werden.





$\bar{n}_{KFA}$ - Abweichungen

$n=10$



Bei den TFA1, IFA und $\overline{SMC}$ - Darstellungen sind die Grenzen für sichere Lösungen mit eingetragen. Sie werden später genauer erläutert.

Die Darstellungen 15 bis 18 zeigen die Abweichungen der jeweiligen faktorenanalytischen Lösungen (Bias) als Flächen im dreidimensionalen Raum. Die Koordinate der mittleren Kommunalität in der Stichprobe $\overline{h}_S^2$ und die Koordinate der Versuchspersonen beschreiben zusammen die $\overline{h}_S^2$ - Ebene. Sie stellt die Ebene der erwarteten Ergebnisse dar. Die dritte Achse beschreibt die Abweichungen von der Erwartung, den systematischen Fehler oder Bias.

1.3.6. Die Bias - Darstellungen 15 bis 18

Die schon in Kapitel IV.1.2. besprochenen Ergebnisse sollen hier nicht wiederholt werden.

Aus diesen Darstellungen sind zusätzliche Feststellungen zu den Grenzwerten für $N \rightarrow \infty$ zu gewinnen. Bei der TFA- Fläche scheint es sich, wenn nur der Bereich der sicheren Lösungen herangezogen wird, um eine fast ebene Fläche zu handeln. Für zunehmendes N läßt sich schließen, daß die Annäherung an die $\overline{h}_S^2$ - Ebene nur sehr langsam erfolgt.

Für IFA, KFA (und damit entsprechend auch für AFA und TFA_{SMC}) sowie für $\overline{SMC}$ läßt sich folgern, daß die bei $N = 200$ erreichten Bias - Werte sich jeweils nur noch geringfügig in ihrem Betrag vergrößern werden (Hyperbeln für feste $\overline{h}_S^2$). Hier handelt es sich schon um gute Schätzungen des Bias für $N \rightarrow \infty$ bei $n = 10$.

1.4. Der systematische Fehler der $\overline{SMC}$

1.4.1. Die Zielsetzung

In den bisherigen Betrachtungen mußte immer wieder hervorgehoben werden, daß die $\overline{SMC}$ - Werte aus zwei Gründen vor allen anderen Analyseergebnissen bevorzugt betrachtet werden müßten:

- a. Unabhängigkeit von der Faktorenanzahl
- b. Hohe Empfindlichkeit bei Vpn - Anzahländerung

Könnte nun die Fläche aus Darstellung 18 ($\overline{SMC}$ - Bias bei $n = 10$) durch eine allgemeine Formel beschrieben werden, so hätte jeder, der sich mit faktorenanalytischen Ergebnissen beschäftigen muß, die Möglichkeit, die gemeinsame Varianz in der Stichprobe zu schätzen. Diese allgemeine Formel müßte allerdings noch um den letzten der 'main effects', die Variablenanzahl, erweitert werden. Damit lautet das Ziel:

A. Bestimmung der Funktion der Fläche aus Darstellung 18

B. Erweiterung dieser Funktion für verschiedene Variablenanzahlen

Die letzte Forderung entspricht der Aufgabe, Variablen'sampling' und Versuchspersonen'sampling' in ihrer jeweiligen Auswirkung auf einen systematischen Fehler zusammen zu erfassen.

Die Möglichkeit, die gemeinsame Varianz in der Stichprobe in Form der mittleren Kommunalität der Stichprobe zu schätzen, entspricht gleichzeitig der Möglichkeit, die Faktorenanzahl genauer zu bestimmen. Die Faktoren dürfen, extrem formuliert, nicht mehr Varianz ausschöpfen, als über solch eine Schätzung der gemeinsamen Varianz beschreibbar ist.

1.4.2. Die allgemeine Formel

Nach vielen Versuchen - und vielen falsch eingeschlagenen Wegen - über die theoretisch vermutete Zerlegung des Bias in zwei Bias - Effekte (s.S.49) gelang es, die in Darstellung 18 vorgestellte $\overline{SMC}$ - Bias - Fläche formelmäßig zu erfassen:

$$(16) \quad \text{Bias}_{\overline{\text{SMC}}} = - \left(\overline{h}_S^2 (1 - \overline{h}_S^2) \right) \frac{N}{N+240} + (1 - \overline{h}_S^2) \frac{5}{N}$$

Um den $\overline{\text{SMC}}$ - Wert insgesamt zu beschreiben, lautet die Formel entsprechend:

$$(17) \quad \overline{\text{SMC}} = \overline{h}_S^2 - \left(\overline{h}_S^2 (1 - \overline{h}_S^2) \right) \frac{N}{N+240} + (1 - \overline{h}_S^2) \frac{5}{N}$$

Eine erste Diskussion dieser Gleichungen läßt folgende Aussagen zu:

Der $\overline{\text{SMC}}$ - Bias ist durch Summation zweier Bias- Effekte gekennzeichnet

$$\begin{aligned} & - \left(\overline{h}_S^2 (1 - \overline{h}_S^2) \right) \frac{N}{N+240} && 1. \text{ Effekt} \\ & + (1 - \overline{h}_S^2) \frac{5}{N} && 2. \text{ Effekt} \end{aligned}$$

1. Effekt :

- A. Für $\overline{h}_S^2 \rightarrow 0$ oder $\overline{h}_S^2 \rightarrow 1$ wird dieser Teil - Bias gegen Null streben. Dies ist durchaus einsichtig. Bei maximalen bzw. minimalen Faktorladungen werden die Variablen kaum noch Kovarianzen mit den Faktoren aufweisen, die über einen SMC - Koeffizienten nicht erfaßt werden.
- B. Für $\overline{h}_S^2 = .50$ strebt dieser Teil - Bias gegen ein Maximum. In diesem Bereich kommen mögliche Kovarianzen der Variablen mit den Faktoren, die durch die restlichen Variablen nicht beschreibbar sind, am stärksten zur Auswirkung. Der $\overline{\text{SMC}}$ unterschätzt die mittlere Kommunalität in der Stichprobe.
- C. Dieser Teil - Bias ist (hier bei $n=10$) von der Versuchspersonenanzahl N abhängig. Für große N strebt er gegen einen Grenzwert, für kleine N gegen ein Minimum.

2. Effekt :

- A. Für $\overline{h}_S^2 \rightarrow 1$ strebt dieser Teil - Bias gegen ein Maximum.

Das bedeutet, daß zufällige Kovarianzen der Variablen sogar im SMC zum Tragen kommen. Im Extrembereich werden sich Fehlerfaktoren im faktorenanalytischen Ergebnis vor den eigentlichen Faktoren darstellen.

- B. Für $\bar{h}_S^2 \rightarrow 1$ strebt dieser Teil - Bias gegen Null. Es kann keine zufälligen Kovarianzen der Variablen mehr geben. Die Varianz der Variablen wird vollständig durch die Faktoren und damit auch durch die SMCs beschrieben.
- C. Auch dieser Teil - Bias ist bei festem $n=10$ von der Versuchspersonenanzahl N abhängig. Für hohe N strebt dieser Teil - Bias gegen Null, bei niedrigem N gegen ein Maximum.

1.4.3. Die allgemeine Formel für beliebige Variablenanzahl

Es geht nun darum, die für $n=10$ beschriebene Formel für beliebiges n zu erweitern. Dabei wird vorausgesetzt, daß sich die Teil - Bias - Ausdrücke in den Klammern nicht verändern werden. Danach muß entsprechend Gleichung (16) eine erweiterte Gleichung in der Form

$$(18) \quad \text{Bias}_{\text{SMC}} = - (\bar{h}_S^2 (1 - \bar{h}_S^2)) K_1 + (1 - \bar{h}_S^2) K_2$$

erwartet werden. K_1 und K_2 sind dabei als Konstanten für feste N und n zu sehen:

$$(19) \quad K_1, K_2 = f (N ; n)$$

Die Aufgabe besteht danach darin, K_1 und K_2 als vom Variablen- und Stichprobenumfang - 'sampling' abhängige Größen zu bestimmen.

Die vorliegenden Simulationsdaten boten Ergebnisse für

$$n = 10, N = 25 / 50 / 100 / 200$$

$$n = 6, N = 200$$

$$n = 14/15, N = 100$$

Sieht man davon ab, daß jeweils Ergebnisse für verschiedene $\bar{h}_S^2$ vorliegen, dann liegt gerade das Minimum von drei Punkten vor,

um den funktionalen Zusammenhang von N und n zu bestimmen. Der erste Lösungsansatz beschränkte sich auch in gewissem Sinn auf diese drei Punkte. Dabei wurde versucht, über eine graphische Lösung (s. Darstellung 6) für die jeweiligen Variablenanzahlen zu bestimmen, bei welcher Versuchspersonenanzahl $\overline{SMC}$ gerade so groß ist wie $\overline{h}_S^2$. Dahinter steht eine Hilfsannahme: Wird bei festem N die Variablenanzahl verändert, dann bedeutet das eine Verschiebung der in Darstellung 18 beschriebenen Fläche. Wenn die Punkte $\overline{SMC} \approx \overline{h}_S^2 = .50$ jeweils für verschiedene Variablenanzahlen bestimmt werden können, dann läßt sich die jeweils benötigte Vpn-Anzahl jeweils bestimmen (Darstellung 6).

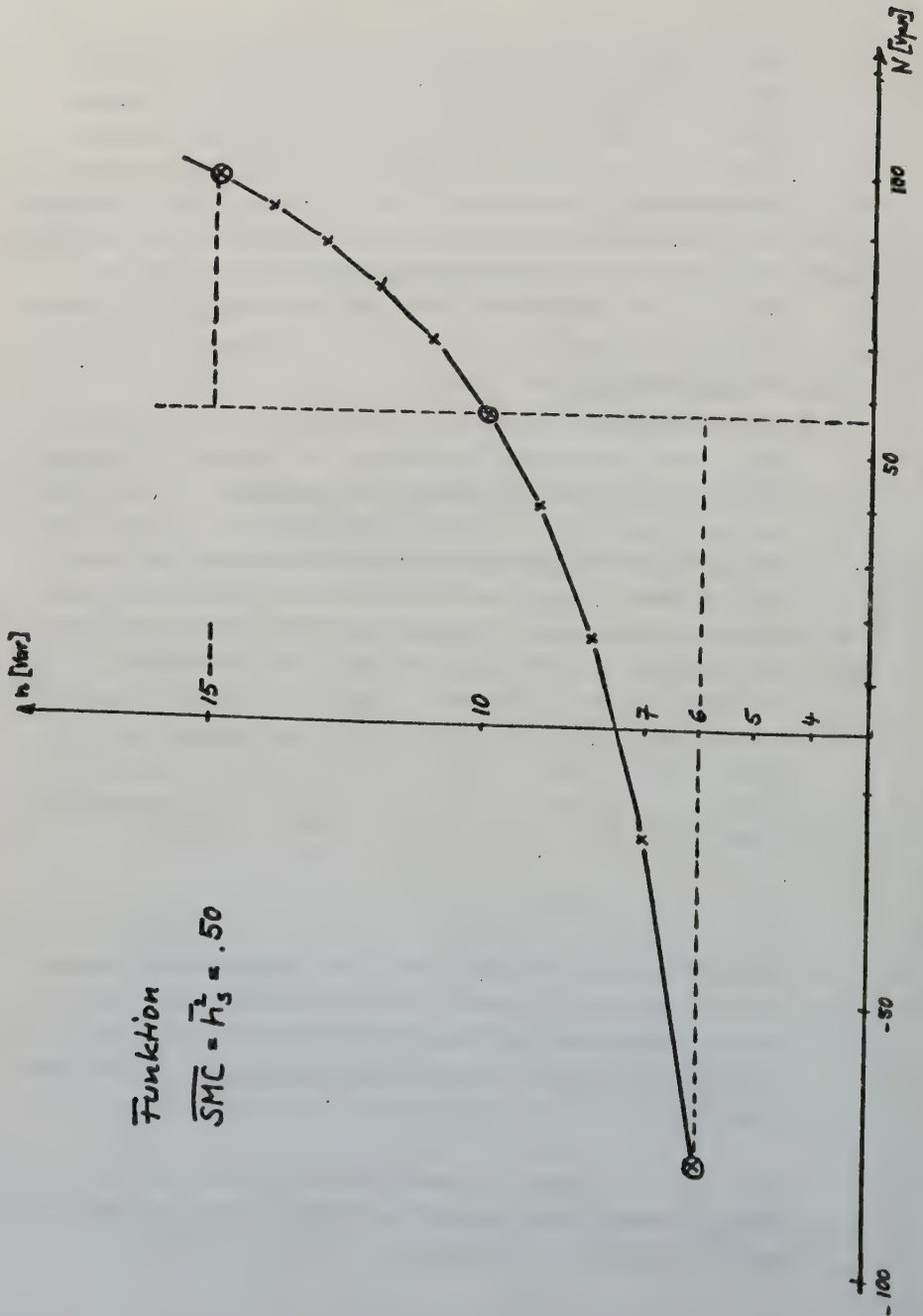
Aus den graphischen Lösungen der Darstellung 6 wurde eine Lösung für K_1 und K_2 erarbeitet. Diese wurden in Gleichung (18) eingesetzt. Bei der Überprüfung der Gleichung an einer Reihe von Ergebnissen aus den Simulationen ergaben sich Unstimmigkeiten. Der Grund dafür lag in der graphischen Lösung der Darstellung 6 für $n=6$: Diese war überlagert von der Hypothese, daß die $\overline{SMC}$ - Koordinate eine Asymptote für die Hyperbel darstellen müsse. Eine neuerliche Überprüfung ergab, daß $\overline{SMC} = \overline{h}_S^2 = .50$ für $n = 6$ im realen Bereich nicht existiert, daß dieser Fall nur für negative Versuchspersonenanzahl beschreibbar ist. Auf dieser Basis wurde neu angesetzt und über jeweilige Überprüfungen allmählich die endgültige Gleichung erarbeitet. Das Ergebnis zeigt Darstellung 19. Die Hyperbelgleichung lautet

$$(20) \quad (N - 160) (n - 3) = - 720$$

In der Darstellung 19 sind die drei oben hervorgehobenen Punkte durch Kreise deutlich gemacht. Es muß allerdings darauf hingewiesen werden, daß Gleichung (20) nicht mehr als Bestimmungsgleichung für diese drei Punkte entstanden ist, sondern über eine Überprüfung anhand der simulierten und analysierten Daten.

Die Gleichung (20) kann in die Gleichungen (16) und (17) als Grundgleichungen eingesetzt werden. Dabei muß nach N (Gleichung (20)) umgeformt werden :

Darstellung 19 : 'Verschiebungsfunktion'



$$N = \frac{-720}{n-3} + 160$$

Der Schnittpunkt der Kurve mit der Geraden $n = 10$ liegt etwa bei $N = 55$:

$$55 = \frac{720}{n-3} - 105 + N$$

Diese Gleichung für $N = 55$ kann in Gleichung (16) oder (17) eingesetzt werden. Dann bestimmen sich K_1 und K_2 als :

$$(21) \quad K_1 = \frac{\frac{720}{n-3} + N - 105}{\frac{720}{n-3} + N + 135}$$

$$(22) \quad K_2 = \frac{5}{\frac{720}{n-3} + N - 105}$$

Geht man von Gleichung (18) in erweiterter Form entsprechend Gleichung (17) aus

$$(23) \quad \overline{SMC} = \overline{h}_S^2 - (\overline{h}_S^2 (1 - \overline{h}_S^2)) K_1 + (1 - \overline{h}_S^2) K_2$$

dann handelt es sich um eine quadratische Gleichung mit einer Unbekannten. Für jede Stichprobe ist $\overline{SMC}$, K_1 und K_2 berechenbar. Löst man nach $\overline{h}_S^2$, der mittleren Kommunalität in der Stichprobe als Unbekannte, die Gleichung (23) auf, dann erhält man

$$(24) \quad \overline{h}_S^2 = \frac{K_1 + K_2 - 1 \pm \sqrt{(1 - K_1 - K_2)^2 - 4 K_1 (K_2 - \overline{SMC})}}{2K_1}$$

oder:

$$(25) \quad \overline{h}_S^2 = \frac{-(1 - K_1 - K_2) \pm \sqrt{(1 - K_1 - K_2)^2 - 4 K_1 (K_2 - \overline{SMC})}}{2K_1}$$

1.4.4. Diskussion der allgemeinen Gleichung

Die folgenden Betrachtungen sollen erhellen, wie sich die erarbeitete Schätzformel für $\bar{h}_S^2$ unter bestimmten Bedingungen verhält. Ausgangspunkt sind die Gleichungen (23) bzw. (25).

A. $\overline{SMC} = \bar{h}_S^2$

Ausgehend von Gleichung (23) ist diese Bedingung erfüllt, wenn die Bias-Effekte in ihrem Betrag gleich groß sind :

$$(\bar{h}_S^2 (1 - \bar{h}_S^2)) K_1 = (1 - \bar{h}_S^2) K_2$$

Diese Gleichung gilt, wenn

a. $(1 - \bar{h}_S^2) = 0$

$$\bar{h}_S^2 = 1$$

b. $(1 - \bar{h}_S^2) \neq 0$

$$\bar{h}_S^2 = \frac{K_2}{K_1} \quad (\text{nach Division durch } 1 - \bar{h}_S^2)$$

$$= \frac{5 \left(\frac{720}{n-3} + N + 135 \right)}{\left(\frac{720}{n-3} + N - 105 \right)^2}$$

Beide Möglichkeiten beschreiben die Schnitte der $\overline{SMC}$ - Fläche mit der $\bar{h}_S^2$ - Ebene. Der Fall a. ist die Gerade $\bar{h}_S^2 = 1$ für alle möglichen N und n. Der Fall b. beschreibt die Schnitt - Hyperbel (s. Darstellung 18), die von N und n abhängt.

Geht, bei festem n, N gegen ein Maximum, dann nähert sich $\bar{h}_S^2$ dem Wert Null. Es handelt sich dabei um die erste Asymptote der Schnitt - Hyperbel. Sie ist Asymptote für alle möglichen n. Die zweite Asymptote wird sich jeweils in Abhängigkeit von der Variablen- und Versuchspersonenanzahl verschieben. Wird n, bei festem N, vergrößert, dann steigt auch der Wert des Verhältnisses K_2/K_1 . Die Bedingung $\overline{SMC} = \bar{h}_S^2$ ist bei gleichem N

und höherem n erst bei höheren Werten von $\bar{h}_S^2$ erfüllt. Das bedeutet, daß die Schnittkurve mit der $\bar{h}_S^2$ - Ebene von höherem n parallel zur V_{pn} - Achse in Richtung eines erhöhten N verschoben ist. Im Hinblick auf die Versuchsplanung bedeutet das : Will man bei einer Erhöhung der Variablenanzahl vergleichbare Schätzungen der $\bar{h}_S^2$ durch die $\bar{SMC}$ erhalten, muß die Versuchspersonenanzahl entsprechend Gleichung (26) erhöht werden.

$$\text{B. } \bar{h}_S^2 = 0$$

In Gleichung (25) kann auf die zweite Lösung (negatives Vorzeichen der Wurzel) verzichtet werden. Diese zweite Lösung beschreibt die negativen $\bar{h}_S^2$ - Lösungen. Sie entstehen dadurch, daß die durch (18) beschriebene Fläche nicht beschränkt wurde auf den Bereich positiver Werte von $\bar{h}_S^2$. Außerdem muß erwartet werden, daß die allgemeine $\bar{SMC}$ - Gleichung (18) eine zweite Fläche beschreibt, die, entsprechend der Ebenendarstellung (Darstellung 6) zu erwartende zweite Hyperbeläste, eine um 180° gedrehte Fläche darstellt. Drehachse ist je nach Variablenanzahl eine in der $\bar{h}_S^2$ - Ebene verlaufende, orthogonal auf der V_{pn} - Achse stehende Gerade bei $N = f(n)$.

Entsprechend dieser Erwartungen im Hinblick auf die durch die allgemeine Formel (18) beschriebenen Bereiche werden sich bei der Kurvendiskussion mehrere Lösungen ergeben. Soweit diese sich auf negative N bzw. $\bar{h}_S^2$ beziehen, sollen sie nicht weiter verfolgt werden.

Um die Lösungen für $\bar{h}_S^2 = 0$ zu beschreiben, kann Gleichung (25) umgeformt werden :

$$(27) \quad \bar{h}_S^2 = \frac{-(1 - K_1 - K_2)}{2 K_1} + \sqrt{\frac{(1 - K_1 - K_2)^2}{4 K_1^2} - \frac{K_2 - \bar{SMC}}{K_1}}$$

$\bar{h}_S^2 = 0$ ist erfüllt, wenn

$$\text{a. } K_1 = \infty$$

Dieser Fall tritt nur für negative N ein, wenn der Nenner von K_1 den Wert Null annimmt.

$$b. K_2 - \overline{SMC} = 0$$

Der Ausdruck unter der Wurzel ist dann das Quadrat des Wertes vor der Wurzel.

$$\begin{aligned}\overline{SMC} &= K_2 \\ &= \frac{5}{\frac{720}{n-3} + N + 135}\end{aligned}$$

Durch diese Gleichung wird die Schnitt-Kurve der $\overline{SMC}$ - Fläche mit der durch die $\overline{SMC}$ - und $\overline{h}_S^2$ - Achse beschriebene Ebene dargestellt. In der Darstellung 18 entspricht dies dem Schnitt der $\overline{SMC}$ - Bias - Fläche mit der durch die Bias - und $\overline{h}_S^2$ - Achsen beschriebenen Ebene. Im Hinblick auf die Simulationen wird hier der $\overline{SMC}$ - Bias bei der Simulation voneinander unabhängiger Variablen beschrieben.

$$c. \overline{h}_S^2 = 1$$

Diese Bedingung ist erfüllt, wenn bei Gleichung (25) S.63 der Zähler so groß ist wie der Nenner:

$$\begin{aligned}-(1-K_1-K_2) + \sqrt{(1-K_1-K_2)^2 - 4K_1(K_2 - \overline{SMC})} &= 2K_1 \\ \sqrt{(1-K_1-K_2)^2 - 4K_1(K_2 - \overline{SMC})} &= 2K_1 + (1-K_1-K_2)\end{aligned}$$

Beide Seiten quadriert, wobei $(1-K_1-K_2) = Z$ gesetzt wird :

$$Z^2 - 4K_1(K_2 - \overline{SMC}) = (2K_1 + Z)^2$$

$$Z^2 - 4K_1(K_2 - \overline{SMC}) = 4K_1^2 + 4K_1Z + Z^2$$

$K_1 = 0$ ist nur für die oben beschriebene zweite $\overline{SMC}$ - Fläche bei größtenteils negativen V_{pn} - Anzahlen zu erwarten. Deshalb kann durch K_1 dividiert werden, ohne für den hier wichtigen Flächenteil zu falschen Ergebnissen zu kommen. Z wird wieder

durch den Klammerausdruck ersetzt:

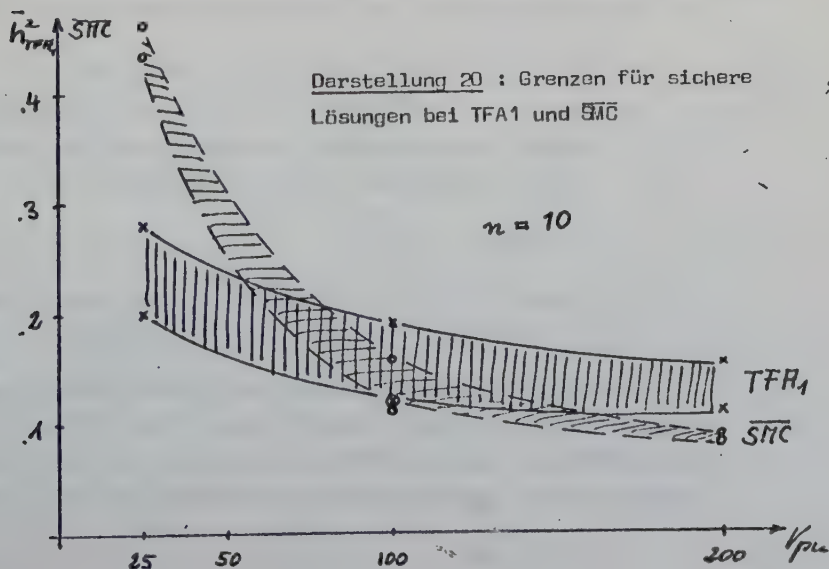
$$-(K_2 - \overline{SMC}) = K_1 + (1 - K_1 - K_2)$$

$$\overline{SMC} = 1$$

Die bedingung $\overline{h}_S^2 = 1$ ist nur dann erfüllt, wenn auch $\overline{SMC} = 1$ ist.

1.4.5. Die Grenze für sichere Lösungen

In den Darstellungen 15 und 18 sind jeweils die Grenzen für sichere Lösungen beschrieben. Diese wurden über eine Ausnutzung von Validitätskoeffizienten (s.S.11) gewonnen. Jedes Analysenergebnis aus den fünf Stichproben je Vpn - Anzahl und je $\overline{h}_S^2$ wurde daraufhin überprüft, ob die simulierten Faktorwerte nach der jeweiligen Faktorenanalyse mit Rotation durch die geschätzten Faktorwerte reproduziert werden. Wichtiger als die Höhe der Validität war dabei die Frage, ob die faktorenanalytischen Ergebnisse die 'eigentlichen' Faktoren an den ersten Stellen der Eigenwertreihe wieder aufzeigen. Als Grenze für sichere Lösungen wird der Bereich betrachtet, in dem sich Fehlerfaktoren vor die eigentlichen Faktoren schieben oder die eigentlichen Faktoren nicht mehr rein, sondern aufgespalten, erscheinen.



In der Darstellung 20 sind jeweils Bereiche angegeben. Der größere Wert beschreibt die erste ungenaue Lösung, der kleinere Wert die Grenze, unter der keine richtige Lösung mehr zu erwarten ist.

Es wäre nun durchaus möglich, die in der Darstellung 20 erhaltenen Hyperbeläste durch entsprechende Gleichungen, die aus den Daten zu entwickeln wären, zu beschreiben. Diese würden den speziellen Fall mit zehn Variablen erfassen. Da bei den vorliegenden Simulationen der Grenzbereich nur durch die Stichproben mit sechs Variablen und mit zehn Variablen erreicht wurde, wäre die Erarbeitung einer allgemeinen Gleichung kaum möglich.

Aus diesen Gründen soll versucht werden, größtenteils deduktiv zu einer allgemeinen Gleichung für den Grenzbereich sicherer Lösungen zu gelangen. Diese kann dann an den vorhandenen Daten überprüft werden. Außerdem soll eine Beschränkung auf die $\overline{SMC}$, aus den bekannten Erwägungen heraus, berücksichtigt werden.

Ausgangspunkt ist Gleichung (23)

$$\overline{SMC} = \underbrace{\overline{h}_S^2 - (\overline{h}_S^2 (1 - \overline{h}_S^2)) K_1}_{I.} + \underbrace{(1 - \overline{h}_S^2) K_2}_{II.}$$

Im Hinblick auf den Grenzbereich für sichere Lösungen können bei den $\overline{SMC}$ zwei Hauptgruppen von Varianzanteilen beschrieben werden :

$$I. \overline{h}_S^2 - (\overline{h}_S^2 (1 - \overline{h}_S^2)) K_1$$

Durch diesen Ausdruck wird der Betrag an der gemeinsamen Varianz, der Varianzanteil am Raum der Faktoren, der durch die $\overline{SMC}$ erfaßt werden kann, beschrieben.

$$II. (1 - \overline{h}_S^2) K_2$$

Hierdurch wird zufällig entstehende gemeinsame Varianz, die auch in den $\overline{SMC}$ zum Tragen kommt, erfaßt.

Auf der Basis dieser Aufteilung kann nun folgende Schlußfolgerung gezogen werden: Wird der II. Anteil größer als der I. Anteil, dann kann erwartet werden, daß sich ein Fehlerfaktor mit solch hohem Gewicht versieht, daß er in der Eigenwertreihe vor einen eigentlichen

Faktor tritt.

Danach gilt für sichere Lösungen:

$$\bar{h}_S^2 - (\bar{h}_S^2 (1 - \bar{h}_S^2)) K_1 > (1 - \bar{h}_S^2) K_2$$

Diese Bedingung kann schärfer gefaßt werden :

$$\bar{h}_S^2 = (1 - \bar{h}_S^2) K_2$$

umgeformt:

$$(28) \quad \bar{h}_S^2 = \frac{K_2}{1 + K_2}$$

(28) beschreibt die Grenze für sichere Lösungen, bezogen auf $\bar{h}_S^2$. Für die $\overline{SMC}$ - Gleichung gilt, daß diese Grenze erreicht ist, wenn die beiden Varianzanteile gleich groß sind :

$$\overline{SMC} = \underbrace{\frac{K_2}{1 + K_2}}_{I.} + \underbrace{\frac{K_2}{1 + K_2}}_{II.}$$

$$\overline{SMC} = \frac{2 K_2}{1 + K_2}$$

Eine Überprüfung dieser Grenze für sichere Lösungen bei den $\overline{SMC}$ anhand der vorliegenden Daten ergab zu geringe Werte im Hinblick auf die ersten auftretenden ungenauen Lösungen. Diesen empirisch gewonnenen Werten genügt die Gleichung

$$(29) \quad \overline{SMC}_{\text{Grenze}} = \frac{\frac{2 K_2}{1 + K_2}}{2}$$

Damit liegt eine Funktion vor, die in Abhängigkeit von der Variablenanzahl und der Versuchspersonenanzahl kritische Werte für die $\overline{SMC}$ angibt. Faktorenanalysen sind nur dann gerechtfertigt, wenn der $\overline{SMC}$ - Wert über dem kritischen Wert liegt. In den sonstigen Fällen sind Fehlerfaktoren in den Ergebnissen unvermeidlich.

2. Der systematische Fehler auf der Einzelvariablenebene

- Der Nivellierungsfehler -

2.1. Vorbemerkungen

Wenn über die mittleren Kommunalitäten von der Stichprobe zur faktorenanalytischen Lösung je Verfahren ein typischer Bias feststellbar ist, muß konsequenterweise auch nach dem Bias auf der Ebene der einzelnen Variablen gefragt werden. Dabei muß vorläufig erwartet werden, daß die einzelnen Kommunalitäten entsprechend dem Bias bei der mittleren Kommunalität verändert werden.

Die Idee, den Validitätskoeffizienten als Korrelationskoeffizient zwischen ursprünglich generierten und nach der Analyse geschätzten Faktorwerten durch eine Korrelation der Faktorladungen zu ersetzen, läßt sich nicht verwirklichen. Hier liegen meist keine gleichen Grundgesamtheiten (Simple Structures) vor. Dagegen könnte eine Korrelationsberechnung zwischen ursprünglichen und aus der Analyse erhaltenen Kommunalitäten durchgeführt werden. In den hier vorliegenden Stichproben handelt es sich aber um maximal 15 Variablen - zu wenig, um auf diesem Weg interpretierbare Ergebnisse zu erhalten. Wenn trotz dieser Einschränkungen die Korrelationen zwischen den h_S^2 und den h_A^2 berechnet wurden, so geschah das in der Absicht, erste Hinweise und Aussagen auch für weitere Analysenschritte zu erhalten. In diesem Sinne werden diese Ergebnisse auch hier dargestellt.

Die weitere Analyse der Daten mußte dann andere Wege einschlagen, um den Veränderungen der h_A^2 im Vergleich zu den h_S^2 gerecht zu werden. Dies ist schon weiter vorne (S.31) beschrieben worden. Jeder h_A^2 - Wert wird Bias - korrigiert und mit dem entsprechenden h_S^2 - Wert verglichen.

Im Hinblick auf die Rotationsergebnisse gelten die bisherigen Einschränkungen : Auch die Kommunalitäten der einzelnen Variablen ändern sich bei der Rotation nicht. Über Veränderungen der Ergebnisse durch bestimmte Rotationsverfahren kann auf diesem Weg keine Aussage gewonnen werden.

2.2. Ein Beispiel einer Stichprobe

Dieses Beispiel ist zur Demonstration besonders geeignet. Es weist über alle Variablen besonders hohe h_S^2 - Abweichungen von $\bar{h}_S^2$ auf.

Kennwerte der Stichprobe:

$$N = 50 ; n = 10 ; k = 2$$

$$\text{Populationsfaktorladungen : } .60 \quad \longleftrightarrow \quad .36$$

(5 Variablen laden den ersten, 5 den zweiten Faktor

mit .60 ; die Vorzeichen der Ladungen wechseln)

$$\bar{h}_p^2 = .49$$

In der ersten Tabelle (Auswertungsbeispiel S.72) sind die jeweiligen h^2 und $\bar{h}^2$ aufgeführt. In der zweiten Tabelle sind für jede Variable i aufgeführt

$$\text{in der ersten Zeile : } h_S^2 - \bar{h}_S^2$$

$$\text{restliche Zeilen : } h_A^2 - h_S^2 - \text{Bias}_{\text{Analyseverfahren}}$$

Beispiel einer Variable :

Die Kommunalität der ersten Variable weicht um $-.071$ von der mittleren Kommunalität in der Stichprobe ab. Bei der TFA1 zeigt sich, daß die Kommunalität dieser Variable nicht wieder im Vergleich zur mittleren Kommunalität niedrig ausgeprägt ist. Bias - bereinigt bewegt sich das h_S^2 dieser Variable um $+.059$ gegen das Mittel. Durch die IFA dagegen wurde die im Vergleich zum Mittel niedrige Kommunalität dieser Variable noch verringert ($-.016$).

Für die Interpretation insgesamt kann vereinfacht folgendermaßen vorgegangen werden: Gleiche Vorzeichen bei den jeweiligen Analyseverfahren wie bei den Abweichungen der Kommunalitäten der Variablen vom $\bar{h}_S^2$ in der Stichprobe zeigt Erhöhung der Abweichung an, entgegengesetzte Vorzeichen bedeuten Verringerung der Abweichung.

In dem gesamten Beispiel mit 50 Kommunalitäten einzelner Variablen aus 5 verschiedenen Analyseverfahren erscheinen 46 Verringerungen und nur 4 Vergrößerungen der Abweichungen vom Mittel. Nach dem Vorzeichentest ist dieses Ergebnis sehr signifikant. Dabei bleibt unberücksichtigt, daß die Vergrößerungen der Abweichungen nur bei

AUSWERTUNGSBEISPIEL ¹⁾

N = 50 , n = 10 , k = 2 , Populationsfaktorlad. .60 .36

A. Kommunalitäten

Var.Nr.	1	2	3	4	5	6	7	8	9	10	$\sum h^2/n$
h_S^2	420	444	735	294	608	338	537	543	684	304	491
h_{TFA1}^2	571	455	678	508	656	629	569	499	680	680	583
h_{IFA}^2	305	347	530	318	413	381	427	376	456	366	392
SMC	395	514	530	408	456	445	442	449	480	414	453
h_{AFA}^2	415	497	653	478	463	446	503	438	535	473	490
h_{KFA}^2 ²⁾	425	493	551	454	448	480	480	484	512	462	479

B. Abweichungen vom Mittel

Var.Nr.	1	2	3	4	5	6	7	8	9	10	$\bar{h}_A^2 - \bar{h}_S^2$
$h_S^2 - \bar{h}_S^2$	-71	-47	+244	-197	+117	-153	+46	+52	+193	-187	
TFA1	+59	-81	-149	+122	-44	+199	-60	-136	-96	+184	+92
IFA	-16	+2	-106	+123	-96	+142	-11	-68	-129	+161	-99
SMC	+13	+108	-167	+152	-114	+145	-57	-56	-166	+148	-38
AFA	-4	+52	-81	+185	-144	+109	-33	-104	-148	+169	-1
KFA	-7	+37	-196	+148	-176	+130	-69	-71	-184	+146	-12

den eher geringen h_S^2 - Abweichungen von $\bar{h}_S^2$ auftreten. Inwiefern die Position der Variablen im Hinblick auch auf den Faktor mit der höchsten extrahierten Varianz eine Rolle spielt, muß zusätzlich geklärt werden.

1) Alle Daten x1000 und gerundet

2) $KFA - h^2$ aus SPSS umgerechnet nach $h_{KFA}^2 = \frac{(h_{SPSS}^2)}{e} : (1 + \frac{(h_{SPSS}^2)}{e})$

2.3. Die Übereinstimmung von Input- und Outputkommunalitäten

Die Darstellungen 21 bis 23 auf den folgenden Seiten zeigen die Korrelationen der h_S^2 mit den jeweiligen h_A^2 über die Variablen.

Bei Darstellung 21 und 22 sind Stichprobenergebnisse verwandt, die den Stichproben Oberlas entsprechen. Im Vergleich zu Oberlas Validitätskoeffizienten (Oberla 1968, S.281) ergeben sich um .200 verminderte Werte.

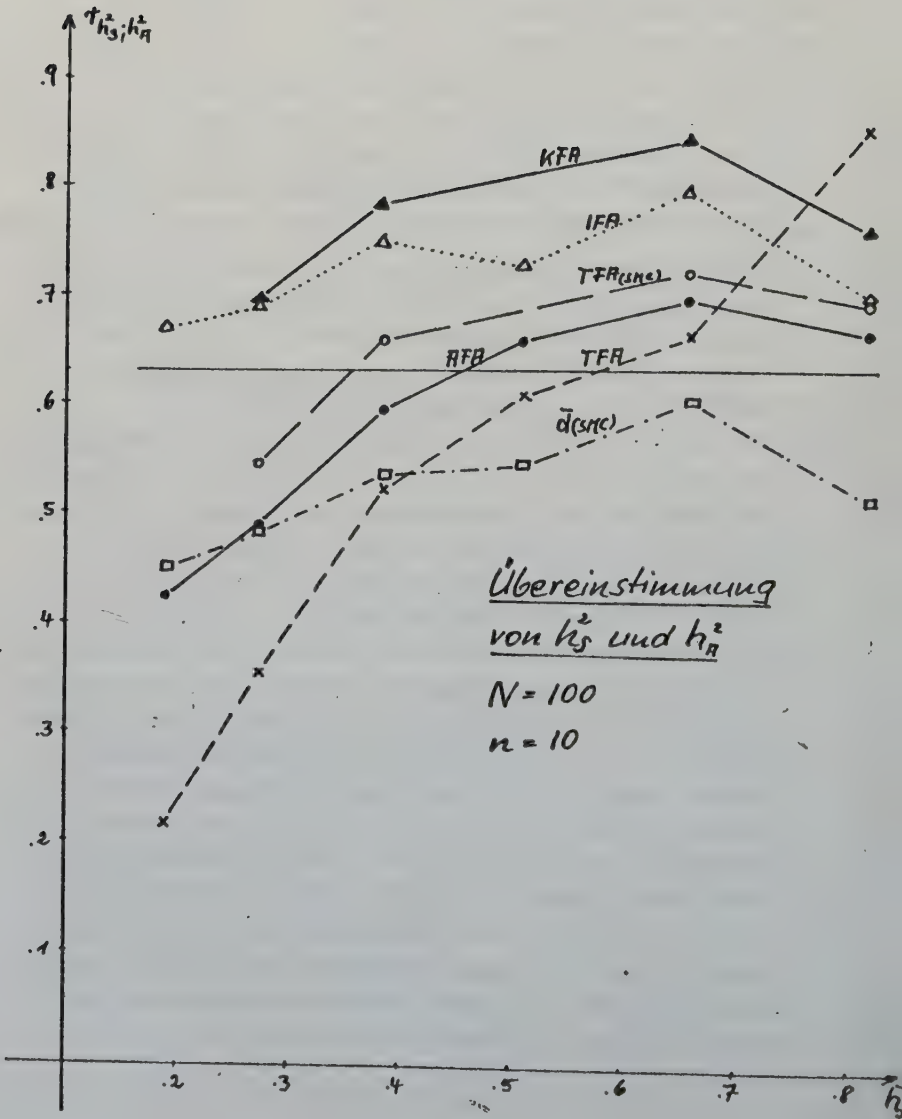
Ein Vergleich der verschiedenen Analysenverfahren läßt sich auf der Basis dieser Korrelationen nicht durchführen. Ebenfalls nicht interpretierbar sind die Differenzen aus den Ergebnissen der Stichproben mit 100 Vpn gegenüber denen mit 200 Vpn. Bei der geringen Anzahl von Variablen sind die Unterschiede nicht signifikant.

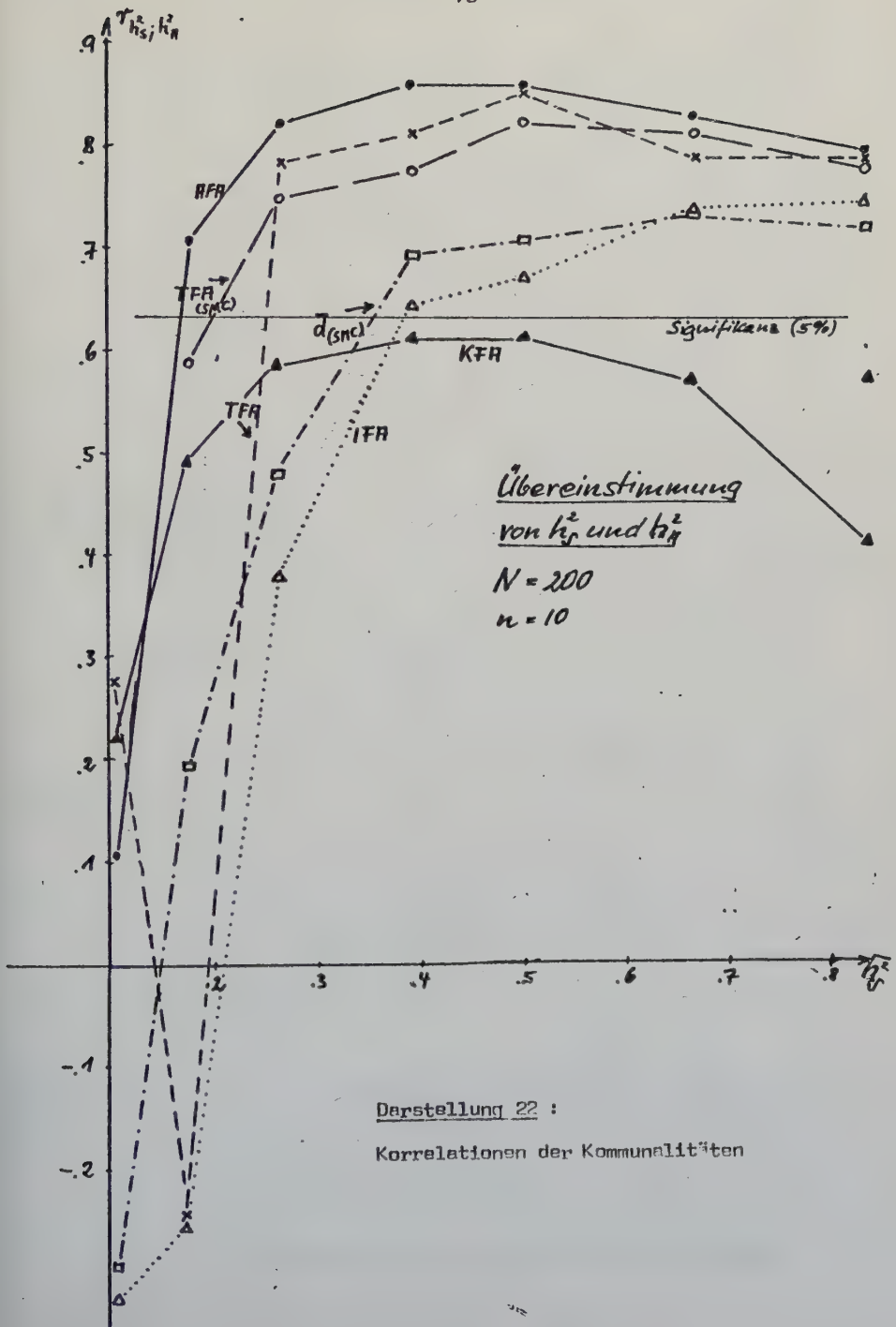
Für die in dieser Untersuchung betonte Bezugnahme auf die SMCs ist bedeutsam, daß diese SMCs sich über alle Bereiche vergleichbar den Ergebnissen aus den einzelnen faktorenanalytischen Verfahren zeigen. Wenn sich bisher Gründe für eine Bevorzugung der SMCs ergaben, dann spricht von diesen Ergebnissen her nichts dagegen.

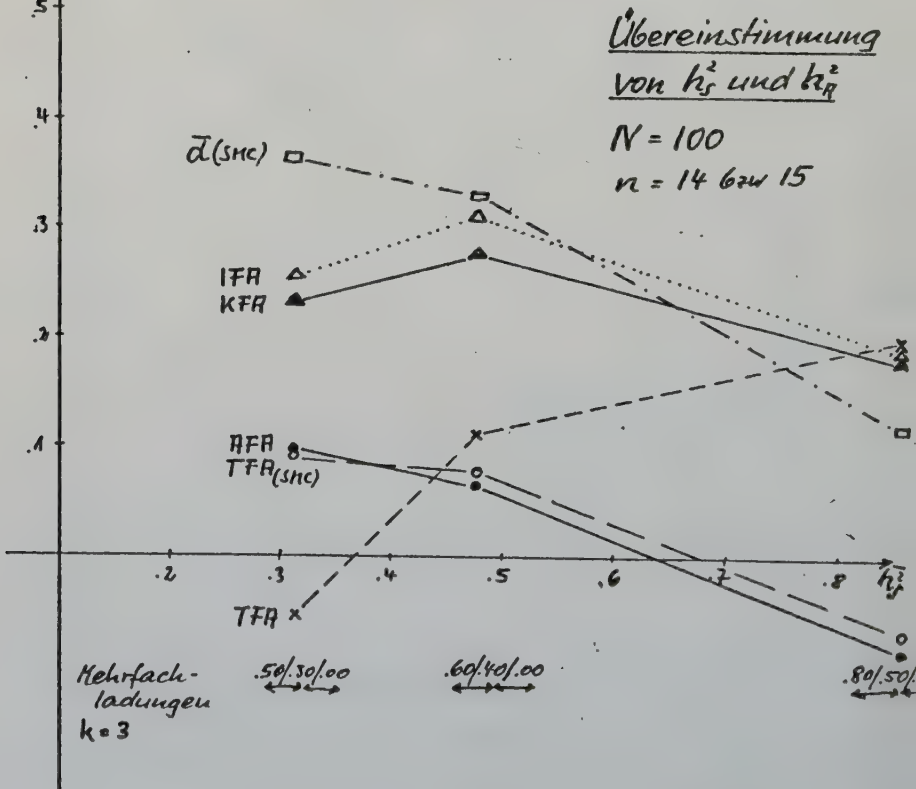
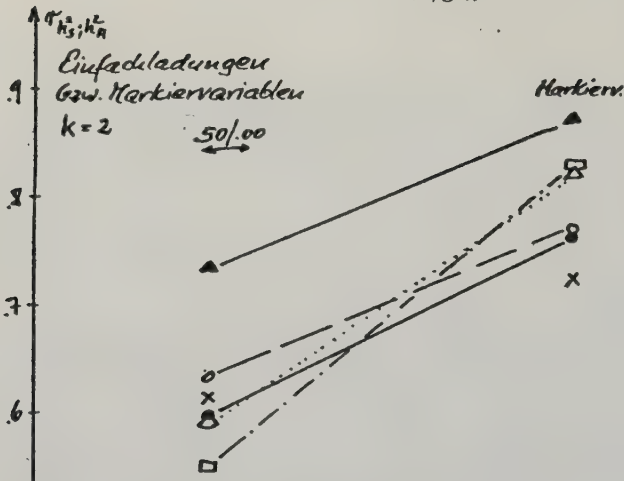
Ein Hinweis für gezielte weitere Betrachtungen sind die Ergebnisse aus der Darstellung 23. Die hohen Korrelationen sind Ergebnisse aus Stichproben mit 14 Variablen und zwei Faktoren. Sie entsprechen den Ergebnissen aus den Stichproben mit 10 Variablen und zwei Faktoren. Die niedrigen Korrelationen zeigen sich bei den Stichproben mit 15 Variablen und drei Faktoren.

Eine erste Vermutung könnte den Unterschied auf die ausgesprochene Verletzung der 'simple structure' - Kriterien bei den Stichproben mit drei Faktoren zurückführen. Dies gilt aber in gewissem Umfang auch für die Stichprobe mit zwei Faktoren und Markiervariablen, besonders wenn nicht die Populationsfaktorladungen, sondern die Stichprobenfaktorladungen betrachtet werden. Damit sind die unterschiedlichen Ergebnisse eher auf die unterschiedliche Faktorenanzahl zurückzuführen. Als Hypothese, unter Einschluß der sich in dem Beispiel aus S.72 abzeichnenden Ergebnisse, könnte formuliert werden : Je höher die Anzahl der Faktoren, um so stärker macht sich eine Nivellierungstendenz bei den Kommunalitäten aus der Analyse bemerkbar.

Darstellung 21 : Korrelationen der Kommunalitäten







Darstellung 23 : Korrelationen der Kommunalitäten

2.4. Der Nivellierungsfehler

Die Darstellungen 24 bis 26 sind drei Stichprobenbeispiele. Jeweils die unterste Zeile zeigt die Abweichungen der h^2 vom Mittel $\bar{h}^2$ in der Stichprobe. Auf einen Vergleich mit der Population wurde verzichtet. Es geht ja darum, die Leistungen der einzelnen Analyseverfahren im Hinblick auf die Input-Daten, die Stichproben-Daten, zu erfassen.

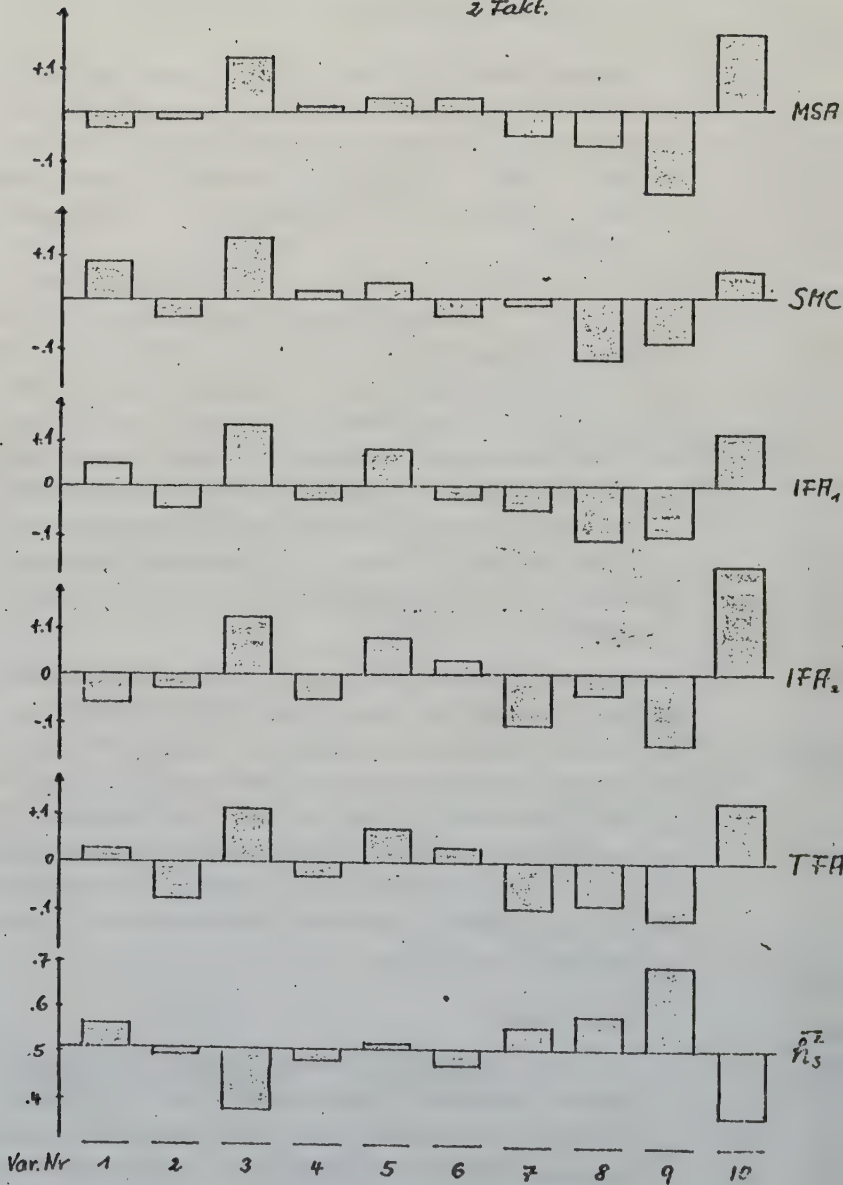
Die Zeilen darüber zeigen jeweils, wie die einzelnen Analyseverfahren die h^2 - Werte der Variablen reproduzieren. Null-Werte bedeuten eine Reproduktion des h_S^2 -Wertes entsprechend seiner über-, unter- oder durch-schnittlichen Ausprägung in der gesamten Stichprobe. Solche Null-Werte über alle Variablen würden bedeuten, daß das Analyseverfahren die Stichprobenkommunalitäten wieder entsprechend darstellt, mit dem Bias versehen. Überdurchschnittliches h_S^2 ist nach der Analyse durch ein überdurchschnittliches h_A^2 beschrieben und interpretierbar. Ein h_S^2 unter dem Durchschnitt erscheint als h_A^2 unter dem Durchschnitt und ist als solches interpretierbar.

Das Auswertungsbeispiel S.72, weitere Auswertungsbeispiele im Anhang und die Darstellungen 24 bis 26 sollen zeigen, inwiefern die einzelnen Analyseverfahren diese Erwartungen von Null-Werten, von identischer Abbildung der h_S^2 , wirklich erfüllen. Es würde hier zu weit führen, über die 11 aufgeführten Stichprobenauswertungsbeispiele hinaus alle 35 auf diesem Weg analysierten Stichproben aufzuführen. Es zeigt sich bei all diesen Stichproben und bei allen jeweils angesetzten Faktorenanalysen (ebenso bei SMC und MSA) dasselbe Bild.

Die Ergebnisse :

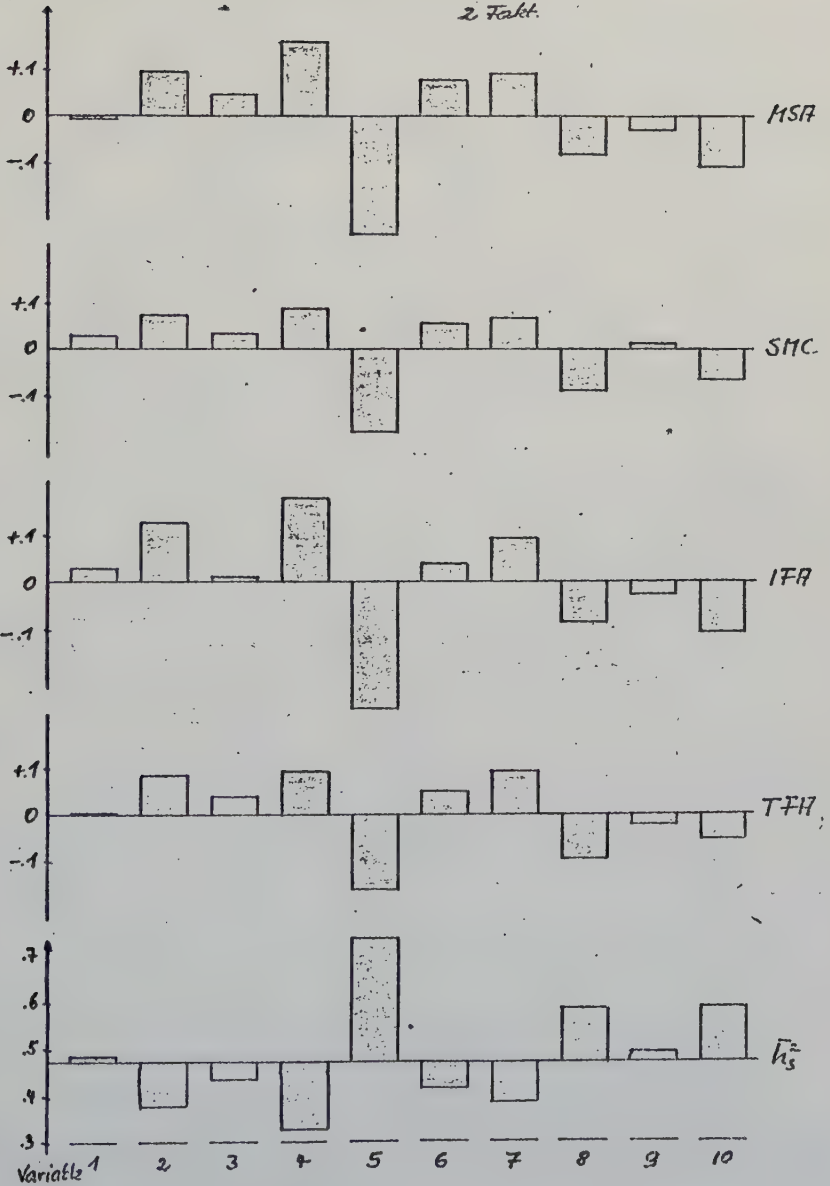
A. Alle in die Untersuchung mit einbezogenen faktorenanalytischen Verfahren zeigen den systematischen Fehler 'Nivellierung' in den Ergebnissen. Dieser Begriff der 'Nivellierung' wurde schon ohne weitere Beschreibung eingeführt. Er soll die Eigenart eines Analyseverfahrens kennzeichnen, im Vergleich zur mittleren Stichprobenkommunalität hohe Input - Kommunalität einer Variablen

Nivellierung 25 Vpu, $h_p^2 = .49$
10 Var
2 Fakt.



Darstellung 24 : 1. Beispiel für den Nivellierungsfehler

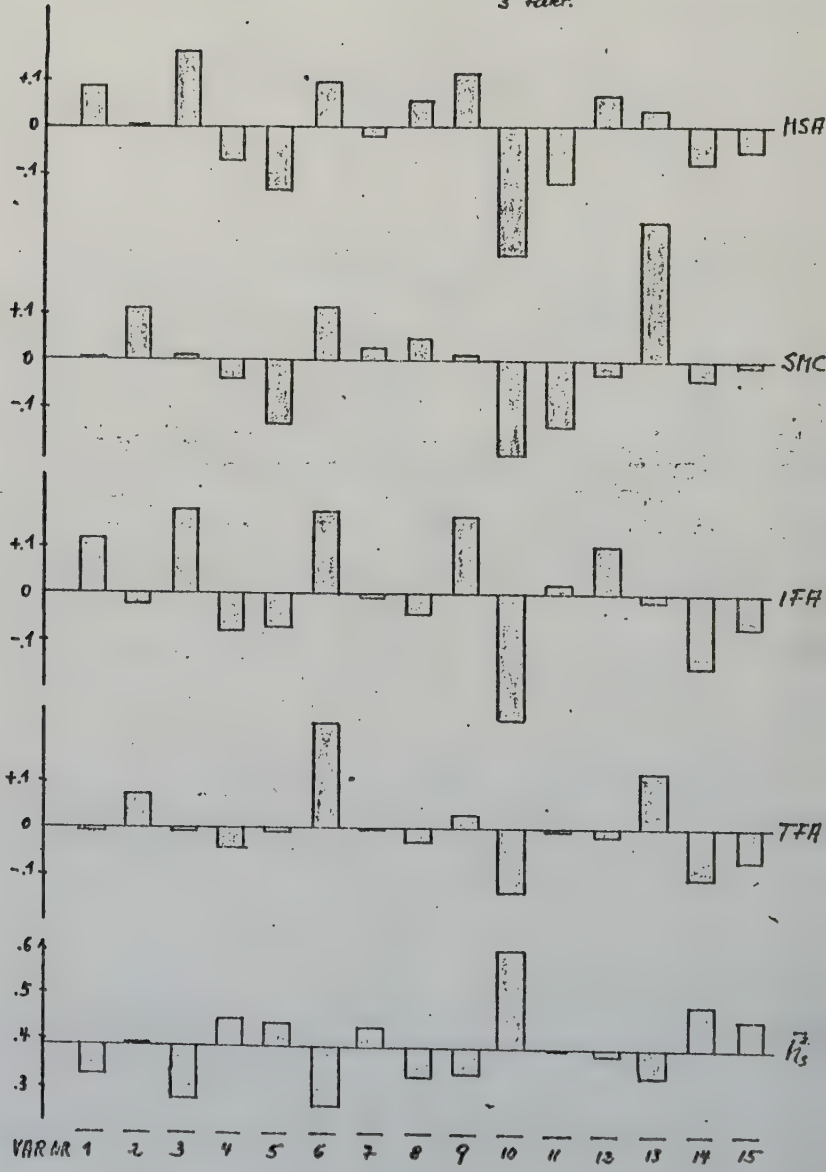
Nivellierung 100 Vm, $k_p = +9$
 10 Var.,
 2 Fakt.



Darstellung 25 : 2. Beispiel für den Nivellierungsfehler

Nivellierung

100 Vpu, Rep.faktor. 50/50/50
15 Var
3 Fakt.



Darstellung 26 : 3. Beispiel für den Nivellierungsfehler

niedriger, niedrige Input- Kommunalität im Vergleich zur mittleren Output - Kommunalität eher höher darzustellen.

- B. Je größer die h_S^2 - Abweichung einer Variablen von $\bar{h}_S^2$ ist, um so größer ist der Betrag der Nivellierung. (s. auch die Darstellungen 27 und 28 sowie die Auswertungsbeispiele im Anhang)
- C. In der Darstellung 24 sind die Ergebnisse aus der Image-Kovarianz-Matrix (IFA_1) und der Image-Korrelations-Matrix (IFA_2) aufgeführt. Beide Analysenergebnisse zeigen die Nivellierung. Dies ist ein Hinweis auf die Unabhängigkeit des Nivellierungsfehlers von der Art der Skalierung der faktorisierten Matrizen. Diese Aussage kann auch aus den im Anhang vorgestellten Auswertungsbeispielen im Vergleich von TFA1 und TFAK gewonnen werden.
- D. Bisher wurden die SMCs global mit unter dem Begriff Analysenverfahren subsumiert. Dies läßt sich für den Bereich des Nivellierungsfehlers weiter beibehalten. Die SMCs sind ebenfalls durch die Nivellierung gekennzeichnet.
- E. Die eingangs durchgeführte kritische Beleuchtung der MSA - Koeffizienten Kaisers erfährt hier eine Bestätigung. Auch diese MSAs mit dem Anspruch, die Angemessenheit des 'Sampling' zu beurteilen, sind der Nivellierung unterworfen.
- F. Ein Teil der Stichprobendaten wurde auf die 'falsche' Anzahl von Faktoren faktorisiert und rotiert. Die Berechnung der h^2 - Werte der Variablen für die zwei 'wahren' Faktoren aus den fünf eingegebenen Faktoren ergab ebenfalls den Nivellierungseffekt.
- G. Die sieben Analysenverfahren (SMC und MSA mit eingerechnet) sind bei allen 35 Stichproben durch den Nivellierungsfehler gekennzeichnet. Dies gilt für alle stärkeren Abweichungen der h_S^2 von $\bar{h}_S^2$.
- F. Die umfangreichen Versuche, den Nivellierungsfehler formal in den Griff zu bekommen, sind bisher gescheitert. Dabei ging es unter anderem darum, das Analysenverfahren zu finden, das diesem zweiten systematischen Fehler am wenigsten unterliegt. Denn in der Praxis ist es im nachhinein nicht mehr bestimmbar, welche Variable in der Stichprobe sehr hohe oder niedrige Kommunalitäten aufwies. Die Probleme bei der Suche nach Möglichkeiten, den Nivellierungs-

fehler auch quantitativ zu beschreiben, zeigen sich unter anderem in folgenden Punkten :

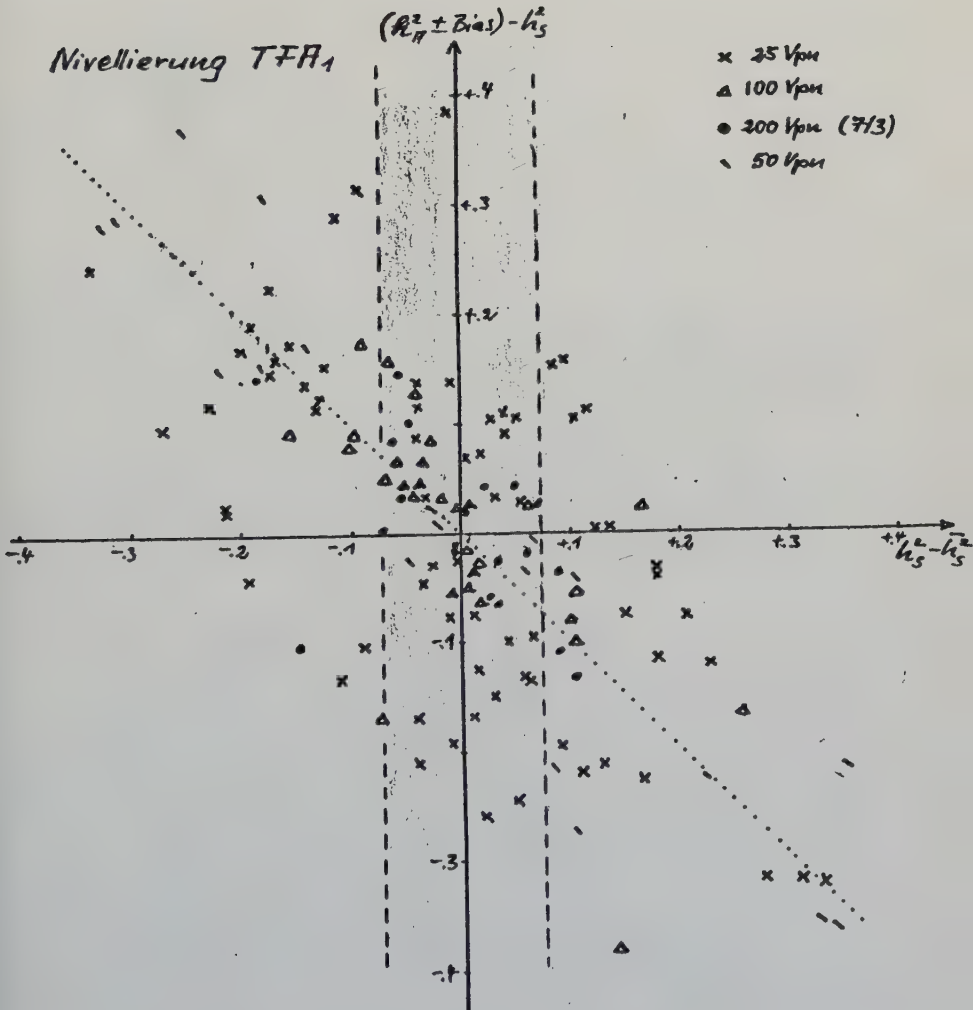
- a. Sollen alle Variablen mit einbezogen werden?
- b. Wie soll das Ausmaß der Nivellierung erfaßt werden? Wenn je Variable bezogen wird auf die h_S^2 - Abweichung von $\bar{h}_S^2$, dann gewinnen die Variablen mit geringer h_S^2 - Abweichung ein zu großes Gewicht. Das gilt vor allem dann für die Fälle, wo das Gegenteil einer Nivellierung auftritt.
- c. Werden nur die bedeutsamen h_S^2 - Abweichungen berücksichtigt (wobei dafür ein Kriterium bestimmt werden muß), dann liegen für zuverlässige Aussagen zu wenig Meßwerte vor.
- d. Das oben angesprochene Kriterium für die Bestimmung extremer Abweichungen ist für sich wieder abhängig von der mittleren Kommunalität. Dabei ergeben sich Probleme im Hinblick auf den 'Deckeneffekt' bei hohen bzw. niedrigen mittleren Stichprobenkommunalitäten. Hier sind extreme Abweichungen nur in einer Richtung möglich.

- I. Die Darstellungen 27 und 28 zeigen den Nivellierungsfehler über einige Stichproben für TFA1 und SMC. Dabei wurde bewußt auf die Stichproben mit $N = 200$ und der Variablenaufteilung 7/3 je Faktor statt 5/5 zurückgegriffen. Die 5/5 Stichproben hätten eine kaum noch darstellbare Verdichtung um die gepunktet gezeichnete 2. Winkelhalbierende gebracht. Diese 2. Winkelhalbierende müßte bei völliger Nivellierung alle Werte auf sich beschreiben.

Ein Vergleich von TFA1 und SMC zeigt, daß die SMC - Werte geringer um die 2. Winkelhalbierende streuen. Das gilt auch noch, wenn ein Bereich geringer h_S^2 - Abweichungen unbeachtet bleibt (Bereich zwischen den gestrichelten Geraden). Die Abbildung der h_S^2 durch die SMCs ist mit größerer Wahrscheinlichkeit durch eine Nivellierung gekennzeichnet. Diese Aussage gilt entsprechend für zunehmende Versuchspersonenanzahlen, sowohl bei TFA1 wie bei den SMCs.

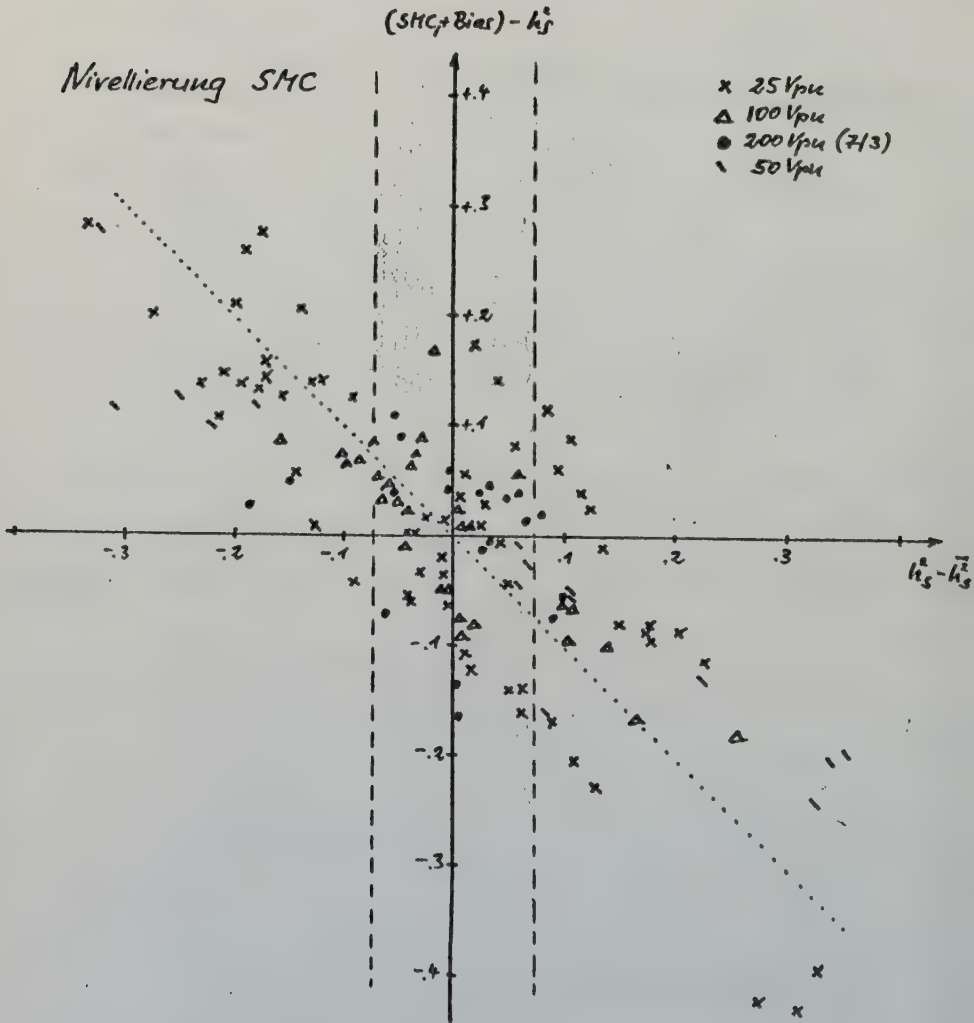
Simulationen mit größerer Variablenanzahl müssen klären, ob für den Nivellierungsfehler bestimmte Gesetzmäßigkeiten bestimmbar sind.

Darstellung 27 : Nivellierung bei der TFA1



Darstellung 28 : Nivellierung bei den SMCs

Nivellierung SMC



2.5. Konsequenzen aus dem Nivellierungsfehler

Wie der Nivellierungsfehler bis jetzt erkannt und beschrieben ist, wird er sich offensichtlich jeder Korrekturmaßnahme entziehen. Es kann dem Praktiker anhand seiner Daten nicht beschrieben werden, welche der Kommunalitäten nivelliert worden ist bzw. welche schon in der Stichprobe nur durchschnittlich war. Entsprechend muß die Technik, einen Faktor in Abhängigkeit von den ihn markierenden Variablen zu beschreiben, in diesem Zusammenhang kritisch gesehen werden. An der Tatsache der 'Markierung' wird sich nichts ändern, dagegen ist die Höhe der Ladung nicht ohne weiteres interpretierbar. Es gibt kein verlässliches Merkmal, das solche Ladungen als 'eigentlich noch höher' oder 'eigentlich noch niedriger' interpretieren ließe.

Eine Klärung ist nur von umfangreichen weiteren Simulationen zu erwarten. Dabei müssen möglicherweise vergleichende Betrachtungen der Faktorladungen der Variablen und ihrer Kommunalitäten herangezogen werden. Es ist nach den im folgenden zu treffenden Folgerungen denkbar, daß sich Bias und Nivellierung auch nach der Ladungsverteilung einer Variable auswirken: Ist eine Variable Markiervariable oder läßt sie alle Faktoren gleichmäßig?

2.6. Eine Ableitung des Nivellierungsfehlers

Dieser Versuch, den Nivellierungsfehler theoretisch zu begründen, wird an der allgemeinen Formel für die $\overline{SMC}$ angesetzt. Dies kann unter anderem damit begründet werden, daß alle Analysenverfahren, die hier verwendet wurden, bis auf die Hauptachsenanalyse TFA1, sich der SMCs als Kommunalitätenschätzung bedienen. Es wird sich zeigen, daß für den ausgeschlossenen Fall der TFA1 ebenfalls eine Herleitung möglich ist.

Ausgehend von der Gleichung (23) für mittlere SMC

$$\overline{SMC} = \overline{h}_S^2 - (\overline{h}_S^2 (1 - \overline{h}_S^2)) K_1 + (1 - \overline{h}_S^2) K_2$$

kann für die einzelne Variable erst einmal entsprechend formuliert werden :

$$SMC = h_S^2 - (h_S^2 (1 - h_S^2)) K_1 + (1 - h_S^2) K_2$$

An und für sich könnte diese Gleichung für die SMCs der Einzelvariablen übernommen werden. Die (für die jeweiligen Stichproben) $\overline{SMC}$ - Konstanten K_1 und K_2 werden übernommen. Bei genauerer Betrachtung des ersten, negativen Bias-Effektes fällt auf, daß der in der $\overline{SMC}$ - Gleichung feststellbare Bezug zur multiplen Regression nicht mehr erhalten ist. Deshalb sollen beide Bias-Effekte noch einmal im Hinblick auf die einzelnen Variablen betrachtet werden.

1. Bias - Effekt : Die hier abzuziehende Varianz ist bisher als die Varianz beschrieben worden, die die Variablen mit den Faktoren, nicht aber mit den restlichen Variablen gemeinsam haben. Auf der Ebene der einzelnen Variablen bedeutet das, daß die Varianz einer Variablen, die sie mit den Faktoren, nicht aber mit den restlichen Variablen gemeinsam hat, im SMC nicht beschrieben werden kann. Der erste Bias-Effekt muß also auch auf der Ebene der einzelnen Variable von dem Varianzbeitrag am Raum der gemeinsamen Varianz der restlichen Variablen abhängen. Als Erwartung beschrieben :

$$E (1. \text{ Bias-Effekt}_{\text{Variable } i}) = f (h_{iS}^2 , \overline{h}_S^2)$$

Je größer h_S^2 einer Variablen im Vergleich zu $\overline{h}_S^2$, desto größer ist die Wahrscheinlichkeit gemeinsamer Varianz von Faktoren und dieser einzelnen Variable, die nicht durch die restlichen Variablen beschrieben werden kann. Umgekehrt ist die Wahrscheinlichkeit von Varianzanteilen nur zwischen einzelner Variable und Faktoren um so geringer, je niedriger h_S^2 im Vergleich zu $\overline{h}_S^2$ ist.

Diese Überlegungen bedeuten eine Nivellierung der SMC im Vergleich zu den h_S^2 . Die durch die $\overline{SMC}$ - Gleichung beschriebene Abbildung der $\overline{h}_S^2$ ist nicht identisch mit der SMC - Abbildung der h_S^2 . Die Formel, die der Erwartung für den ersten Bias-Effekt genügt, lautet :

$$1. \text{ Bias-Effekt}_{\text{Einzelvariable}} = (h_S^2 (1 - \overline{h}_S^2)) K_1$$

2. Bias-Effekt : Hier handelt es sich um zufällige gemeinsame Varianz. Sie ist auf der Ebene der einzelnen Variablen abhängig von der Uniqueness der jeweiligen Variable. Die Wahrscheinlichkeit zufälliger gemeinsamer Varianz mit den restlichen Variablen (nicht nur mit deren unigen Anteilen!) nimmt zu mit zunehmendem unigen Anteil der einzelnen Variable, nimmt ab mit abnehmendem unigen Anteil dieser Variable. Auch dieser 2.Bias-Effekt bedeutet auf der Ebene der einzelnen Variablen im Vergleich zu der Gesamtstichprobenebene eine Nivellierungstendenz. Die Abbildungen der einzelnen Variablen im SMC erfolgen nicht parallel der $\overline{SMC}$ - Abbildung. Damit bleibt der 2.Bias-Effekt auf der Ebene der einzelnen Variable als

$$2.Bias-Effekt_{\text{Einzelvariable}} = (1 - h_S^2) K_2$$

Die SMCs der einzelnen Variablen sind daraus beschreibbar als

$$(30) \quad SMC = h_S^2 - (h_S^2 (1 - \overline{h}_S^2)) K_1 + (1 - h_S^2) K_2$$

Der 1.Bias-Effekt ist als SMC-typisch anzusehen. Dagegen kann der 2.Bias-Effekt durchaus auch für die Hauptkomponentenanalyse TFA1 gelten, allerdings mit einer anderen Konstante. Aus dieser Überlegung ergibt sich

$$(31) \quad \overline{h}_{TFA1}^2 = \overline{h}_S^2 + (1 - \overline{h}_S^2) K_3$$

und auf der Ebene der einzelnen Variable

$$(32) \quad h_{TFA1}^2 = h_S^2 + (1 - h_S^2) K_3$$

Der vom Stichprobenumfang und der Variablenanzahl abhängige Koeffizient K_3 ist in seinem Wert sicher nicht identisch mit K_2 . Letzten Endes muß K_3 im Gegensatz zu K_2 auch von der Faktorenanzahl abhängig sein. Je mehr Faktoren extrahiert werden, um so höher wird der Betrag der erfaßten zufälligen gemeinsamen Varianz

anzusetzen sein.

Vorläufig bedeutsamer ist, daß die Überlegungen zum 2.Bias-Effekt bei den SMCs auf der Ebene der einzelnen Variablen auch für die h^2_{TFA1} gelten. Bei der Hauptkomponentenanalyse werden die h^2_{TFA1} ebenfalls nicht wie die $\bar{h}^2$ abgebildet, sondern es kommt zu einer Nivellierung.

Die Ergebnisse aus den Überlegungen zu SMC und h^2_{TFA1} sind auf jede in der vorliegenden Untersuchung eingesetzte Analysenform übertragbar. Bei TFAK, AFA und KFA werden die SMCs der einzelnen Variablen als Kommunalitätenschätzungen verwandt. Sie sind durch die Nivellierung gekennzeichnet und müssen diese in diesen Analysen auch zur Wirkung bringen. Die IFA benutzt ebenfalls die SMCs als Kommunalitätenschätzung. Außerdem sind die nichtdiagonalen Elemente der Image-Kovarianzmatrix über die multiple Regression gewonnen. Beides wird die oben geschilderten Folgen des 1.Bias-Effekt auf der Ebene der Einzelvariable verstärken.

V. Ein neues Verfahren zur Bestimmung der Faktorenanzahl

1. Vorbemerkungen

Es lassen sich bei den praktischen Verfahren zur Bestimmung der Faktorenanzahl eine Reihe von Techniken aufzeigen, die alle auf der Basis folgender Dateninformationen bzw. einer Kombination dieser beruhen :

- A. Absolute Varianzanteile der Faktoren
- B. Relative Varianzanteile der Faktoren an der Gesamtvarianz
- C. Relative Varianzanteile der Faktoren am Raum der gemeinsamen Varianz

Mit der Betonung der 'praktischen Verfahren' sind hier Fragen in Bezug auf den dahinter stehenden Ansatz (Algebraisch, statistisch oder psychometrisch (Oberla 1968, S.123)) oder Probleme der Signifikanzprüfung nicht berücksichtigt.

Kennzeichnend für die in der Literatur vorgeschlagenen praktischen 'Schnell' - Verfahren zur Bestimmung der Anzahl der Faktoren ist, daß sie die Tatsache bzw. Möglichkeit zufälliger gemeinsamer Varianz kaum berücksichtigen.

Oberla (1968, S.127/128) zeigt die Eigenwertreihe einer Korrelationsmatrix aus Zufallszahlen. Dieser stellt er die Eigenwertreihe einer Korrelationsmatrix des 16-Personality-Factor-Questionnaires (vergleichbare Vpn- Und Variablenanzahlen) gegenüber : "Wenn keine Zufallskorrelationen vorliegen, findet sich eine (Eigenwert-) Kurve wie ...". Hier kommt Oberla zu einer irreführenden Aussage, die kurz danach durch die Beschreibung der Horn'schen (1965) Technik der Faktorenanzahlbestimmung indirekt wieder aufgehoben wird. Denn Horns Vorschlag basiert gerade auf der Erkenntnis, daß jede Eigenwertreihe auch durch zufällige gemeinsame Varianz beeinflusst ist.

Im Hinblick auf die Möglichkeiten aus der vorliegenden Untersuchung ist Horns Technik der Faktorenanzahlbestimmung die bedeutsamste :

- a. Bestimmung der Eigenwertreihe einer Zufallszahlen-Korrelationsmatrix mit den Kennwerten der zu analysierenden Stichprobe (N,n)
- b. Eigenwertreihe der zu analysierenden Stichproben - Korrelationsmatrix
- c. Subtraktion der zufälligen gemeinsamen Varianz, die sich bei
a. in Form von Eigenwertreihen größer 1 gezeigt hat, von den entsprechenden Eigenwerten aus b.
- d. Resultierende Eigenwerte größer 1 sind als Eigenwerte 'echter' Faktoren anzusprechen.

Kritisch anzumerken ist, daß Horn wohl die Stichprobenparameter N und n berücksichtigt, nicht aber die Höhe der gemeinsamen Varianz in der Stichprobe, zum Beispiel in der Form einer mittleren Kommunalität. Je höher der Anteil der Kommunalitäten der Variablen gegenüber den unigen Anteilen wird, um so geringer muß die zufällige gemeinsame Varianz ausfallen.

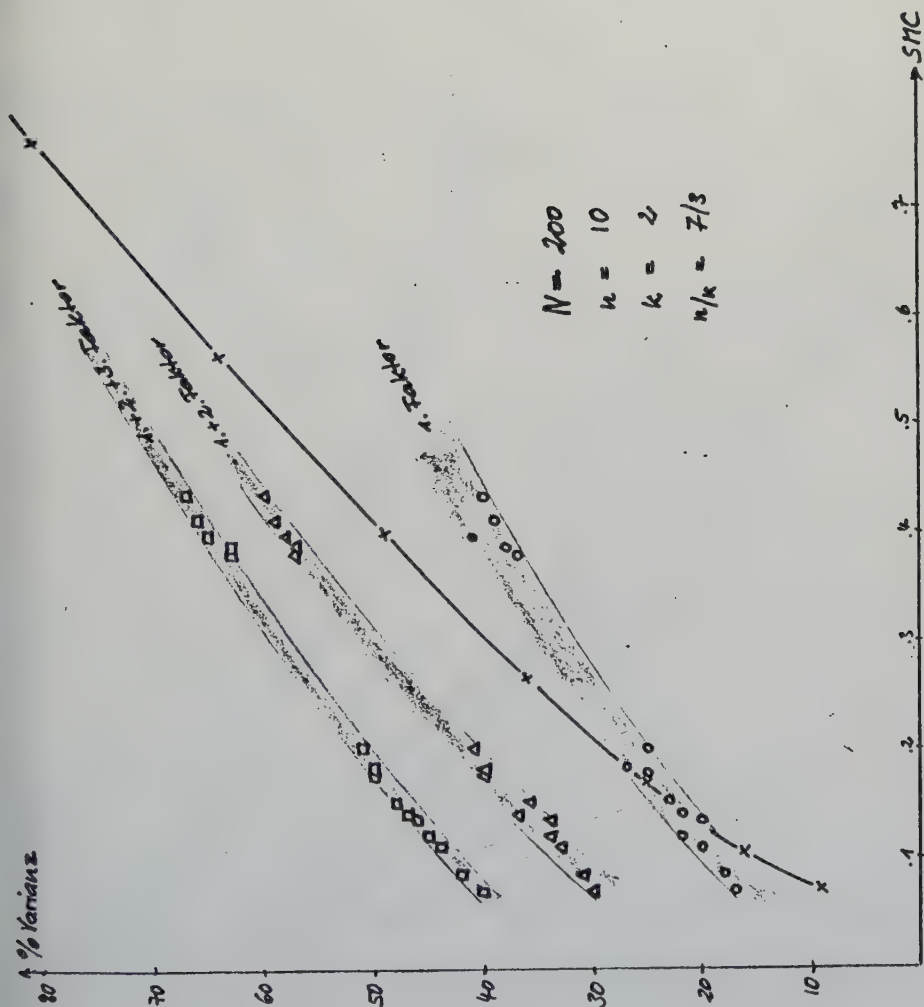
Gegenüber Horns Technik bietet sich nun aus den vorliegenden Ergebnissen die Möglichkeit, über zwei zusätzliche Informationen die Bestimmung der Anzahl der Faktoren zu verbessern :

1. Bestimmung der zufälligen gemeinsamen Varianz in Abhängigkeit von der mittleren Uniqueness
2. Schätzung der mittleren Kommunalität in der Stichprobe

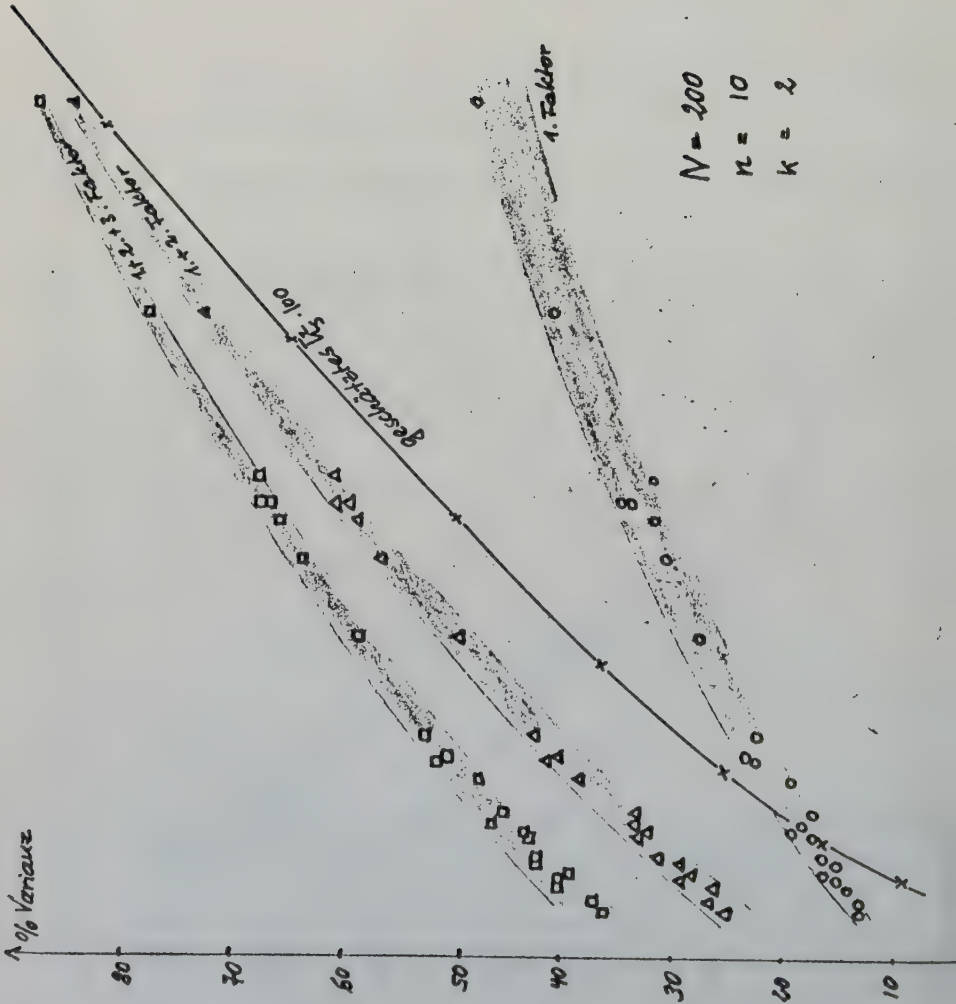
2. Das Verfahren

Nach den bisherigen Ergebnissen muß sich die Faktorenanzahlbestimmung über die $\bar{h}_S^2$ - Schätzung durchführen lassen (Gleichung 25). Diese Schätzung ist direkt als eine Schätzung des Varianzanteiles der 'wahren' Faktoren zu sehen. Ein Vergleich mit der Eigenwertreihe aus der Hauptkomponentenanalyse bzw. den daraus berechneten Prozentanteilen an der Gesamtvarianz müßte die zu berücksichtigenden Faktoren aufzeigen. Allerdings muß dabei mit in Betracht gezogen werden, daß die TFA1 immer zu hohe Varianzanteile (s. Darstellung 15) liefert.

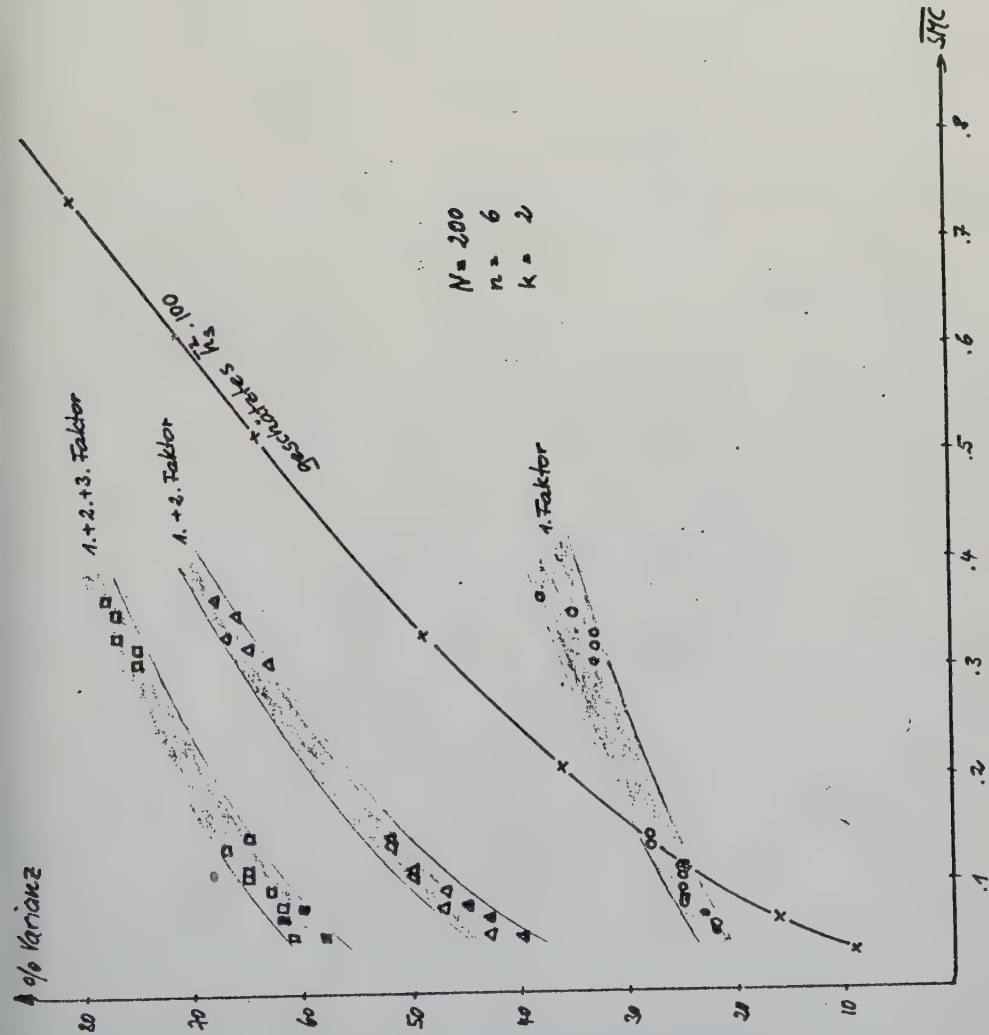
Darstellung 29 : Geschätzte mittlere Kommunität in der Stichprobe
und Gesamtvarianzanteile der Faktoren aus der TFA1



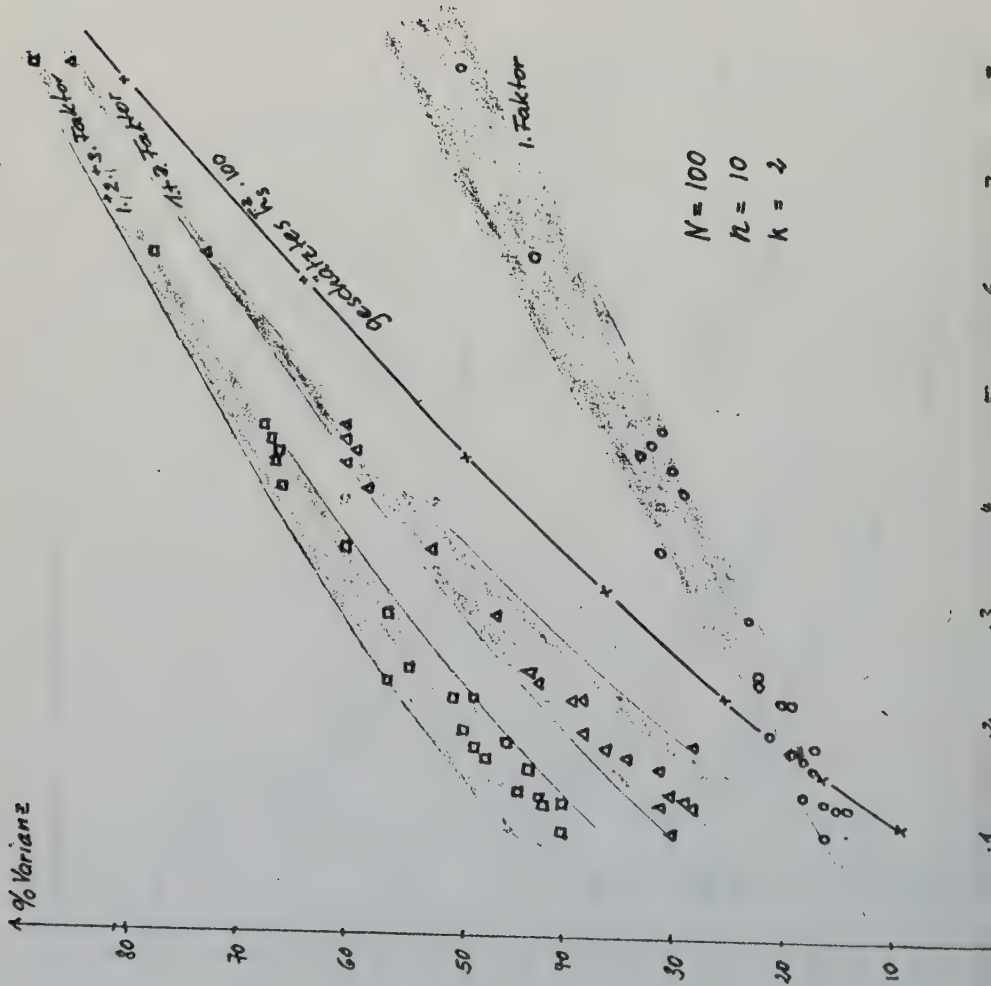
Darstellung 30 : Geschätzte mittlere Kommunalität in der Stichprobe
und Gesamtvarianzanteile der Faktoren aus der TFA1



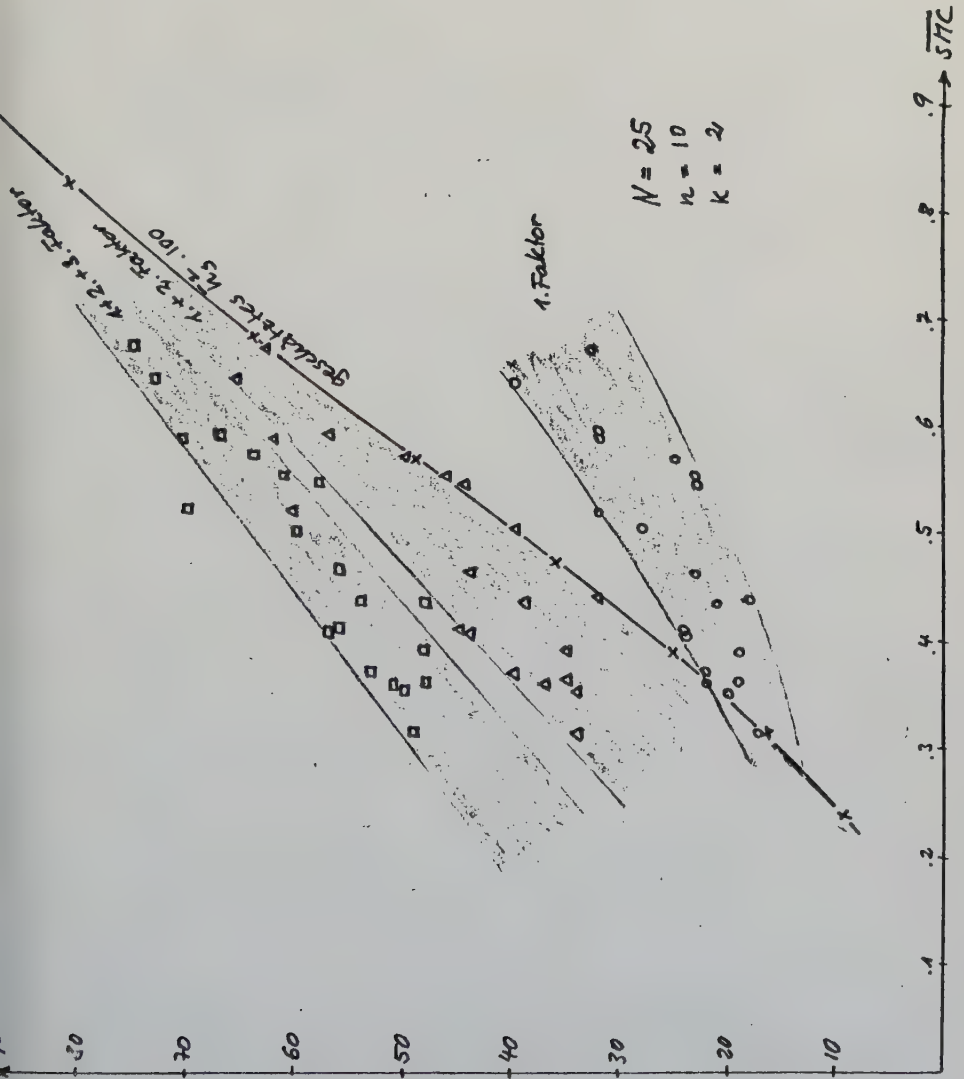
Darstellung 31 : Geschätzte mittlere Kommunalität in der Stichprobe
und Gesamtvarianzanteile der Faktoren aus der TFA1



Darstellung 32 : Geschätzte mittlere Kommunalität in der Stichprobe
und Gesamtvarianzanteile der Faktoren aus der TFA1



Darstellung 33 : Geschätzte mittlere Kommunalität in der Stichprobe
und Gesamtvarianzanteile der Faktoren aus der TFA1



Die Darstellungen 29 bis 33 zeigen die Ergebnisse für jeweils feste N , n und k . Es sind die Summen der prozentualen Varianzanteile der Faktoren und die Kurve der geschätzten mittleren Kommunalität in der Stichprobe in Abhängigkeit von dem jeweiligen $\overline{SMC}$ - Wert aufgeführt. Eine einfache Konsequenz aus den Darstellungen wäre die Empfehlung, die Faktorenanzahl so zu bestimmen, daß deren Varianzanteilsomme die geschätzte gemeinsame Varianz gerade übersteigt.

Daß dieser Ansatz nicht für alle möglichen Bereiche hält, kann aus zwei Hinweisen abgeleitet werden :

- a. Bei den Darstellungen 29 bis 33 übersteigt der Varianzanteil des ersten Faktors den geschätzten Wert $\overline{h}_S^2 \cdot 100$ im Bereich niedriger $\overline{SMC}$, die nicht alle unter der Grenze für sichere Lösungen liegen. Hier ist der Betrag der zufälligen gemeinsamen Varianz, der durch den ersten Faktor mit beschrieben wird, im Verhältnis zu groß.
- b. Bei der Darstellung 14 und deren Diskussion ergab sich schon, daß der bei der TFA1 zu erwartende Bias sich nach der Anzahl der extrahierten Faktoren richtet.

Die daraus resultierende, bisher möglichst vermiedene Konsequenz ist, vor der endgültigen Bestimmung der Faktorenanzahl sich mit der möglichen Anzahl von Faktoren zu beschäftigen. Es wird sich zeigen, daß trotzdem ein 'blindes' Verfahren erhalten bleibt. Es geht darum, die in der TFA1 durch die jeweiligen Faktoren zusätzlich erfaßte zufällige Varianz zu schätzen. Dazu gibt es aus den bisherigen Betrachtungen der Ergebnisse eine Reihe von Erkenntnissen :

- A. Die extrahierte zufällige gemeinsame Varianz ist von $(1 - \overline{h}_S^2)$, der mittleren Uniqueness einer Stichprobe abhängig. (Darstellung 15 und Seite 87)
- B. Je mehr Faktoren bei fester Variablenanzahl in der TFA1 extrahiert werden, um so größer wird der insgesamt zusätzlich eingebrachte zufällige Varianzanteil.
(Verlauf der Eigenwertreihen aus Korrelationsmatrizen aus Zufallszahlen / Darstellung 14)

C. Aus den Darstellungen 30 und 31 läßt sich ableiten :

Je mehr Variablen eine Stichprobe bei gleicher Vpn -Anzahl und gleicher Faktorenanzahl hat, um so geringer wird die durch die Faktoren zusätzlich ausgeschöpfte zufällige gemeinsame Varianz.

Diese einzelnen Punkte führen zu dem Versuch, in Gleichung (31) das K_3 zu bestimmen. Die den simulierten und hauptkomponenten-analysierten Daten am ehesten entsprechende Lösung ist

$$K_3 = \frac{k}{n} \quad \begin{array}{l} k : \text{Anzahl der extrahierten Faktoren} \\ n : \text{Anzahl der Variablen} \end{array}$$

Danach läßt sich Gleichung (31) schreiben als

$$(33) \quad \bar{h}_{TFA1}^2 = \bar{h}_S^2 + (1 - \bar{h}_S^2) \frac{k}{n}$$

In diese Gleichung ist die Annahme eingegangen, daß der Bias bei der TFA1 so zu beschreiben ist, daß jeder zusätzliche Faktor einen gleichen Betrag an Bias-Varianz einbringt.

Die Entwicklung der Gleichung (33) war im Hinblick auf die Darstellungen 29 bis 33 und eine daraus abzuleitende Faktorenanzahlbestimmung notwendig geworden. Danach gilt jetzt: Die aus der TFA1 bestimmte extrahierte Varianz der Faktoren, vermindert um den erwarteten Bias, muß mit dem geschätzten $\bar{h}_S^2$ übereinstimmen.

Die Gleichung (25) könnte in Gleichung (33) eingesetzt werden. Eine Auflösung nach K_3 würde die Möglichkeit eröffnen, auf rein formalem Weg die Anzahl der Faktoren zu bestimmen. Da aber nur in den wenigsten Fällen ganzzahlige Werte für K_3 zu erwarten sind, soll diesem Verfahren ein graphisches Verfahren vorgezogen werden. Dazu wird Gleichung (33) etwas abgewandelt :

$$(34) \quad \bar{h}_{TFA1}^2 = \bar{h}_S^2 + (1 - \bar{h}_S^2) \frac{k}{n+1}$$

Durch diese Korrektur muß die Summe der % der Spur aus der TFA1 vermindert um den Bias $(1 - \bar{h}_S^2) \frac{k}{n+1}$, den aus den $\bar{S}MC$ geschätzten Wert $\bar{h}_S^2 \cdot 100$ gerade übertreffen. Dieser graphische Lösungsweg

kann auch in kritischen Fällen die Möglichkeit eröffnen, den Zuwachs an 'echter' gemeinsamer Varianz, auch im Vergleich zu der zufälligen gemeinsamen Varianz, zu beurteilen.

In diesen Möglichkeiten liegt unter anderem der Grund, die Eigenwerte aus der Hauptkomponentenanalyse für den Vergleich mit der geschätzten mittleren Kommunalität heranzuziehen. Im Sinne von Kaisers II. Prinzip (Kaiser 1970, S.404) ist dabei der 'Simplicity' Aspekt berücksichtigt:

- a. Die zur Schätzung der $\bar{h}_S^2$ verwendeten SMCs sind nicht gleichzeitig in dem Analysenverfahren eingesetzt
- b. Es gibt nur Überschätzungen der $\bar{h}_S^2$ durch die $\bar{h}_{TFA1}^2$. Dieser Bias ist im Vergleich zum Bias bei IFA, TFAK, AFA und KFA einfach zu beschreiben. Es gibt keine Überlagerung durch den Bias der Kommunalitätenschätzung dieser Verfahren
- c. Der TFA1 - Bias ist praktisch nicht durch die Versuchspersonenanzahl berührt. (Darst. 15, 29 und 30)

Damit ist der Weg des Einsatzes faktorenanalytischer Verfahren mit der Bestimmung der Faktorenanzahl festgelegt :

1. Schritt: Hauptkomponentenanalyse der Korrelationsmatrix mit Berechnung der multiplen Korrelationen (SMC)
2. Schritt: Berechnung, ob die Grenze für sichere Lösungen überschritten ist, nach

$$\bar{SMC}_{Limes} = \frac{\sqrt{2K_2 / (1 + K_2)}}{2}$$

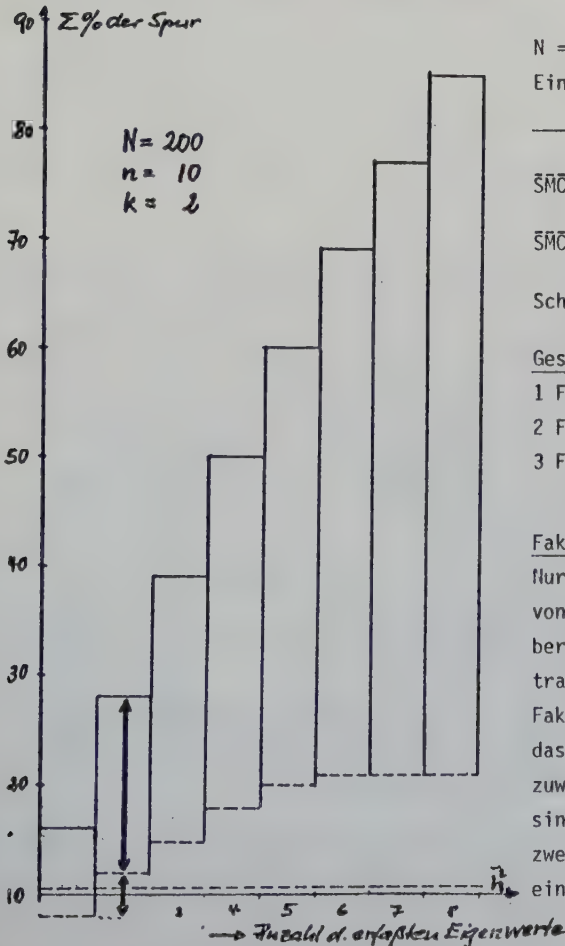
3. Schritt: Schätzung der mittleren Kommunalität $\bar{h}_S^2$ der Stichprobe nach Gleichung (24) oder (25)
4. Schritt: Graphische Darstellung der Summenkurve aus den Prozents der Spur (TFA1) und der Kurve der jeweils subtrahierten $(1 - \bar{h}_S^2) \cdot k/n \cdot 100$ als geschätztem jeweiligem Gesamt-Bias. Der erste Bias-bereinigte Wert, der die geschätzte mittlere Kommunalität übersteigt, kennzeichnet die Faktorenanzahl.

Bei nicht eindeutigen Fällen kann das Verhältnis des Zuwachses an 'echter', d.h. bias-korrigierter, gemein-

samer Varianz zu dem Gesamt-Bias beurteilt werden. Ein Richtwert ist, daß der 'echte' Varianzzuwachs mindestens die Hälfte des jeweiligen gesamten Bias ausmachen sollte.

5. Schritt: Nach den vorliegenden Ergebnissen sollte eines der Verfahren, TFAK, AFA oder KFA bei der endgültigen Analyse verwandt werden. Diese sind durch die geringsten systematischen Fehler gekennzeichnet. Die TFA1 überschätzt, die IFA unterschätzt die Kommunalitäten der Variablen in der Stichprobe.

Ein Beispiel aus den simulierten Daten :



$N = 200$, $n = 10$, $k = 2$
Einfachladungen .30

$\bar{S}\bar{M}\bar{C} = .075$

$\bar{S}\bar{M}\bar{C}_{\text{Limes}} = .075$

Schätzung: $\bar{h}_S^2 = .104$

Gesamt-Bias (in der Darst. $\times 100$)

1 Faktor: $(1 - .104) \cdot \frac{1}{11} = .081$

2 Faktoren: $(.896)^2 / 11 = .162$

3 Faktoren: $(.896)^3 / 11 = .243$

usw.

Faktorenanzahl:

Nur nach der Überschreitung von $\bar{h}_S^2$ durch die bias-bereinigte Summenkurve betrachtet ergeben sich zwei Faktoren. Sowohl $\bar{S}\bar{M}\bar{C}_{\text{Limes}}$ wie das Verhältnis von Varianzzuwachs zu Gesamtbias (Pfeile) sind Hinweise darauf, daß der zweite Faktor möglicherweise ein Fehlerfaktor ist.

Die Überprüfung dieser Aussage anhand der Validitätskoeffizienten (s.S.11) ergibt :

- TFA1 : Der erste Faktor wird korrekt wiedergegeben, während der zweite Faktor erst an dritter Stelle beschrieben wird. Ein Fehlerfaktor hat mehr Varianzanteile als einer der simulierten Faktoren.
- IFA : Der erste simulierte Faktor wird durch die Analyse nur aufgespalten in zwei Faktoren beschrieben. Der zweite Faktor erscheint an der richtigen Position, der Validitätskoeffizient ist aber vergleichsweise niedrig.

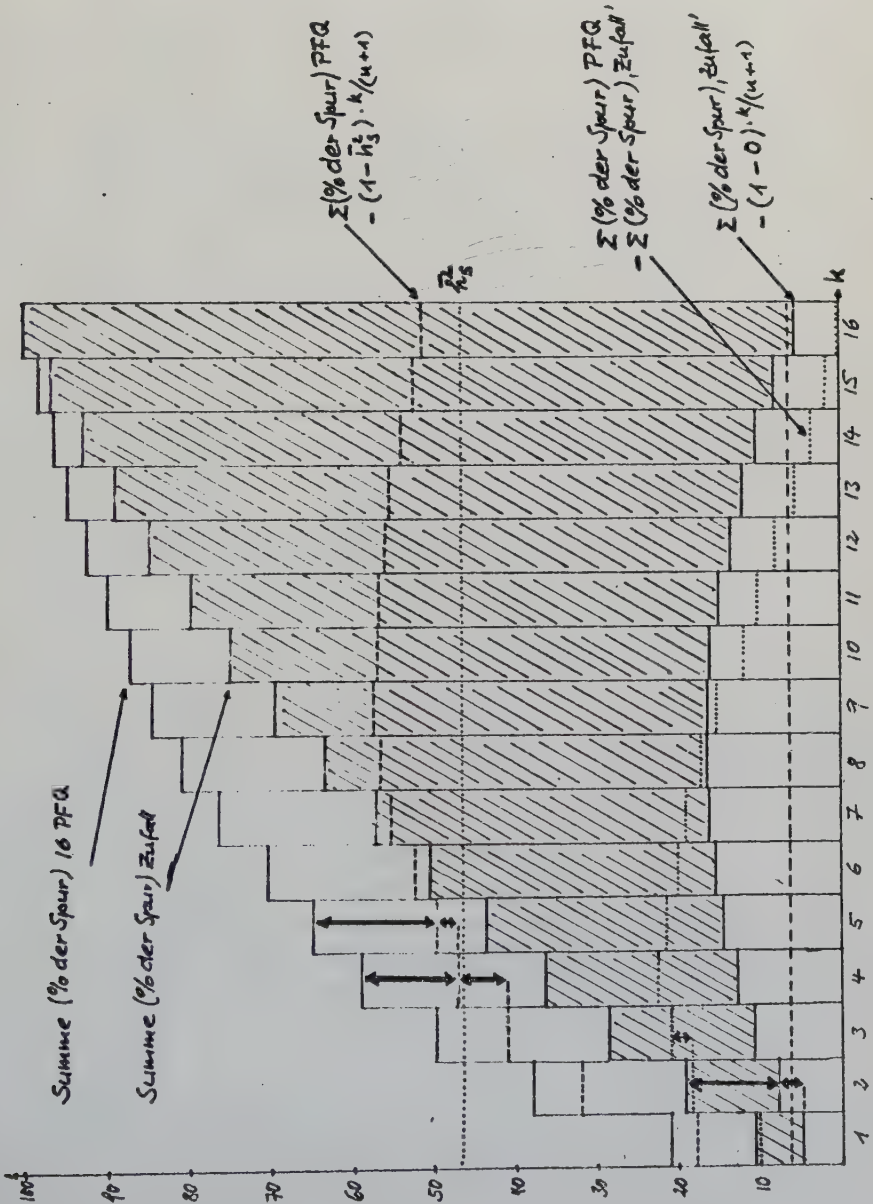
(ZUSATZBEMERKUNG: Dieses Ergebnis der IFA scheint typisch für diese Analysentechnik zu sein. Bei der Betrachtung des Ansatzes der IFA ist auch zu erwarten, daß dieses Verfahren das Vordringen von Fehlerfaktoren eher verhindert. Die früher gehegte Erwartung, daß die IFA insgesamt der TFA1 überlegen sei (Schiller, Wittman 1971) bestätigt sich nicht. Das gilt auch für diesen Grenzbereich. Wenn die TFA1 Fehlerfaktoren vorrücken läßt, dann erscheinen bei der IFA die Faktoren gespalten.)

Ein weiteres Beispiel zu der Korrelationsmatrix aus Zufallszahlen

Zu diesem Beispiel liegen keine eigenen simulierten Daten vor. Dargestellt werden die bei Oberla (1968, S.127/128) beschriebenen Eigenwertreihen einer Korrelationsmatrix aus Zufallszahlen und einer Korrelationsmatrix des 16-Personality-Factor-Questionnaires. Die Variablenanzahlen und die Vpn-Anzahlen sind identisch. In der Darstellung 34 (folgende Seite) wird jeweils die gesamte Eigenwertreihe berücksichtigt :

- Summenkurve (% der Spur) aus Zufallsdaten
- Summenkurve (% der Spur) aus PFQ
- Jeweils die Summenkurve nach Subtraktion von $(1 - \bar{h}_S^2) \cdot k / (n+1)$
($\bar{h}_S^2 = .00$ bzw. $= .47$)
- Die aus dem Horn - Vorschlag (1965) resultierende Kurve aus der Subtraktion der jeweiligen Summenkurvenwerte 'Zufall' von den Summenkurvenwerten PFQ.

Darstellung 34 : Eigenwertreihen in Abhängigkeit von der mittleren Kommunalität in der Stichprobe (.00 bzw. .47)



A. Ergebnisse zu der Korrelationsmatrix aus Zufallszahlen

Ein Versuch, die Faktorenanzahl zu bestimmen, trifft auf sieben Eigenwerte größer Eins. Nach dem Scree - Test könnten zwei Faktoren bestimmt werden. Nach dem Abzug der erwarteten zufälligen Varianz $(1 - \bar{h}_S^2) \cdot k/(n+1)$ ergibt sich, daß der Zuwachs an 'echter' Varianz durch den zweiten Faktor im Vergleich zu der zufälligen Varianz (s.Doppelpfeile im zweiten Feld der Darstellung) äußerst gering ist. Außerdem ist schon durch den ersten Faktor der erwartete $\bar{h}_S^2 = .00$ überschritten. Nach der Bias-Korrektur wird durch den ersten Faktor nicht einmal der Varianzanteil einer Variablen beschrieben.

B. Ergebnisse zu der Korrelationsmatrix des PFQ

Bei vier Eigenwerten größer Eins bestimmt auch der Scree-Test vier Faktoren. Die bias-bereinigte Summenkurve überschreitet den geschätzten Wert $\bar{h}_S^2 = .47$ nur knapp. Das Verhältnis von Zuwachs an 'echter' gemeinsamer Varianz zu gesamter zufälliger Varianz läßt einen vierten Faktor zu. Die Frage, ob möglicherweise erst mit einem fünften Faktor (bias-bereinigt) der $\bar{h}_S^2$ - Wert überschritten wird, kann verneint werden. Der Zuwachs an echter gemeinsamer Varianz ist zu gering.

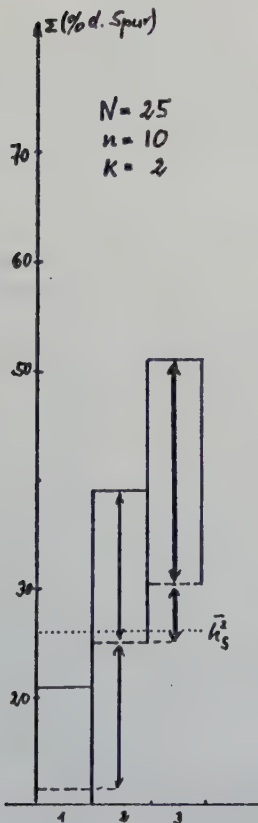
C. Schätzung der zufälligen gemeinsamen Varianz nach Horn

Der Horn - Vorschlag geht darauf hinaus, die 'echte' gemeinsame Varianz dadurch zu schätzen, daß jeweils von der Summenkurve (% der Spur) einer Korrelationsmatrix die Summen (% der Spur) einer entsprechenden Korrelationsmatrix aus Zufallszahlen abgezogen werden. Bei dem vorliegenden Beispiel des PFQ hieße das, daß nur zwei interpretierbare Faktoren vorhanden sind. Der Zuwachs an echter gemeinsamer Varianz durch den dritten Faktor ist zu gering. Dieses Ergebnis ist erwartet worden. Mit Horns Technik wird die zufällige gemeinsame Varianz überschätzt.

VI. ANWENDUNGEN

1. Anwendungsbeispiele aus den simulierten Daten

Die folgenden Beispiele beschränken sich größtenteils auf den ersten, bedeutsamen Teil der Summen der % der Spur. Außerdem sind vor allem Beispiele von Stichproben aus dem 'kritischen' Bereich ausgewählt, um die Stärke der Leistung des entwickelten neuen Verfahrens zur Bestimmung der Faktorenanzahl zu belegen.



BEISPIEL 1

$N = 25$, $n = 10$, $k = 2$

Populationsfaktorladungen (Einfach) .50

$\overline{SMC} = .41$

$\overline{SMC}_{Limes} = .300$

Schätzung: $\overline{h}_S^2 = .26 = 26\%$

Gesamt - Bias

1 Faktor: $(1 - .26) \cdot 1/11 = .067 = 6,7\%$

2 Faktoren: $(1 - .26) \cdot 2/11 = .134 = 13,4\%$

3 Faktoren: $(1 - .26) \cdot 3/11 = .201 = 20,1\%$

Faktorenanzahl :

Bias - bereinigt überschreitet erst eine dreifaktorielle Lösung das geschätzte $\overline{h}_S^2$. Ein Vergleich des jeweiligen Varianzzuwachses im Hinblick auf den Gesamtbias läßt nur eine Lösung mit zwei Faktoren zu.

BEISPIEL 2

$N = 25$, $n = 10$, $k = 2$

Populationsfaktorladungen (einfach) .70

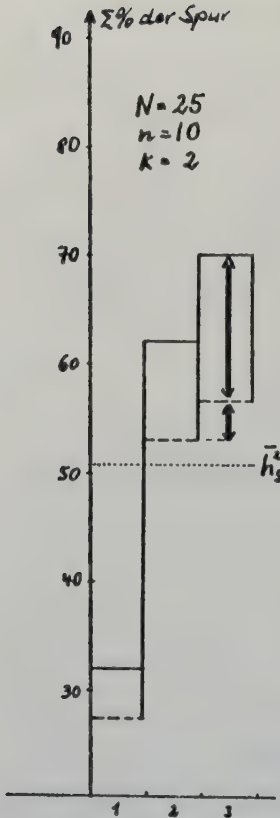
$$\bar{S}\bar{M}\bar{C} = .58$$

$$\bar{S}\bar{M}\bar{C}_{\text{Limes}} = .30$$

$$\text{Schätzung: } \bar{h}_S^2 = .51 = 51\%$$

Faktorenanzahl:

Die bias-bereinigte Summenkurve übersteigt die geschätzte gemeinsame Varianz mit dem zweiten Faktor. Bei einem dritten Faktor wäre der Zuwachs an echter gemeinsamer Varianz zu gering.



BEISPIEL 3

$N = 25$, $n = 10$, $k = 2$

Populationsfaktorladungen (einfach) .30

$$\bar{S}\bar{M}\bar{C} = .316$$

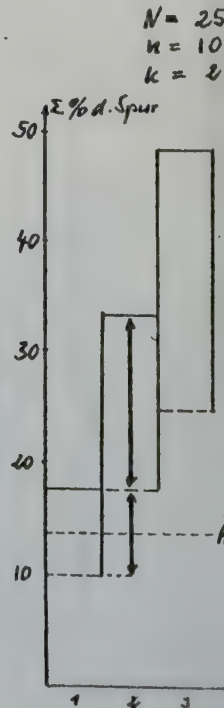
$$\bar{S}\bar{M}\bar{C}_{\text{Limes}} = .30$$

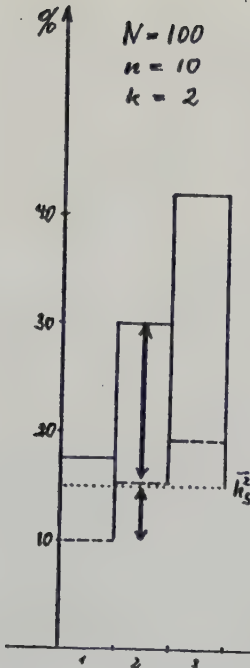
$$\text{Schätzung: } \bar{h}_S^2 = .137$$

Bias- bereinigt übersteigt die Summenkurve mit dem 2. Faktor die $\bar{h}_S^2$ -Schätzung. Der Zuwachs an Varianz ist aber im Verhältnis zu der zufälligen Varianz gering.

Vergleich mit Validitätskoeffizienten:

Nur der erste Faktor wird reproduziert





BEISPIEL 4

$N = 100$, $n = 10$, $k = 2$

Populationsfaktorladungen (einfach) .40

$$\bar{S}\bar{M}\bar{C} = .15$$

$$\bar{S}\bar{M}\bar{C}_{\text{Limes}} = .13$$

Vier Eigenwerte größer 1

$$\text{Schätzung: } \bar{h}_S^2 = .15$$

Faktorenanzahl:

Bias-bereinigt wird mit dem 2. Faktor gerade der geschätzte Varianzbeitrag beschrieben. Das Verhältnis von Zuwachs durch den 2. Faktor zu der Bias-Varianz läßt keine genaue Lösung zu.

Vergleich mit den Validitätskoeffizienten: Die ursprünglich simulierten zwei Faktoren werden nach der Analyse durch die ersten vier Faktoren beschrieben.

BEISPIEL 5

$N = 100$, $n = 10$, $k = 2$

Populationsfaktorladungen (einfach) .40

$$\bar{S}\bar{M}\bar{C} = .17$$

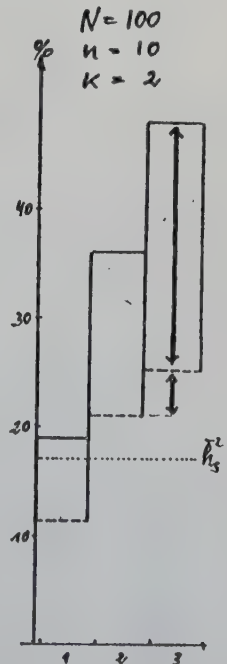
Drei Eigenwerte größer 1

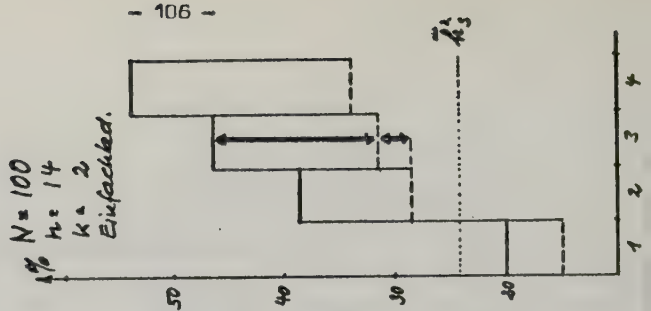
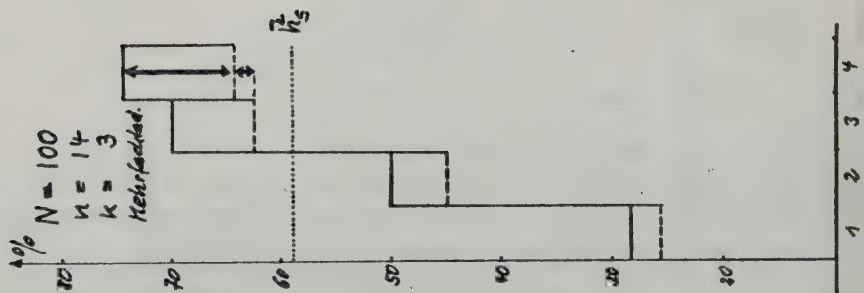
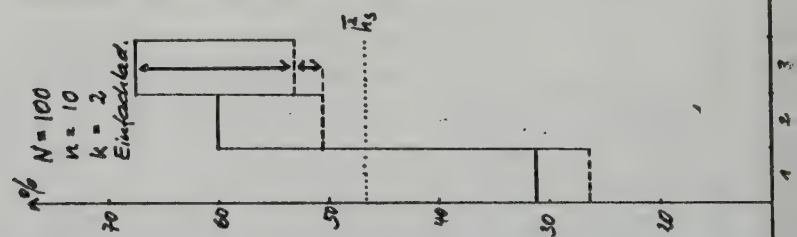
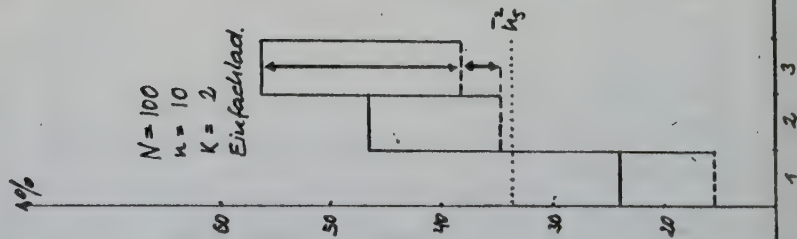
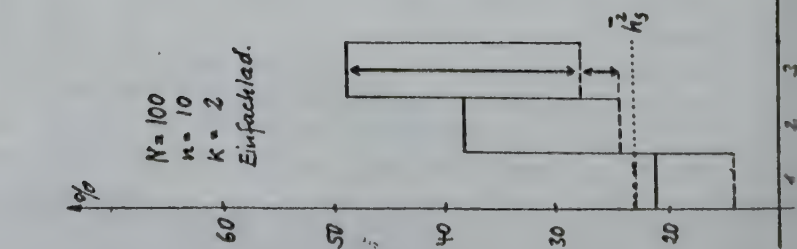
$$\text{Schätzung: } \bar{h}_S^2 = .17$$

Faktorenanzahl:

Die bias-bereinigte Varianzsumme übersteigt mit dem zweiten Faktor die geschätzte gemeinsame Varianz. Erst bei dem dritten Faktor ist der Varianz-Zuwachs zu gering.

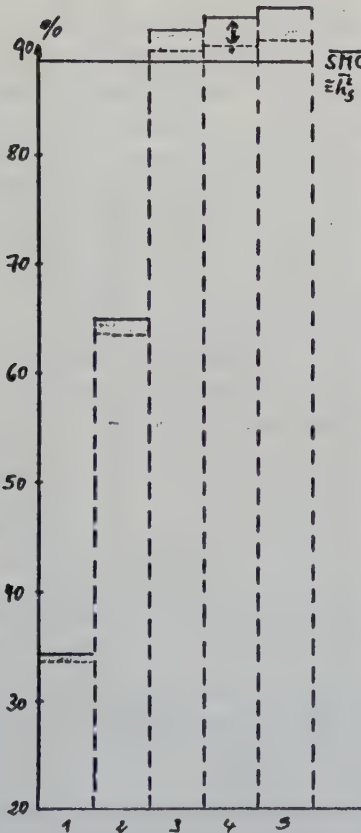
Auch die Validitätskoeffizienten belegen, daß die ursprünglichen Faktoren reproduziert wurden.





Für die Beispiele 6 bis 10 erübrigt sich eine ausführliche Dateninformation. Die Bestimmung der Faktorenanzahl ist auf diesem Weg eindeutig. Das gilt auch für die Beispiele 6 und 10, bei denen sich jeweils vier Eigenwerte größer 1 ergeben hatten.

Das letzte Beispiel aus der Reihe der simulierten Stichproben soll noch einmal demonstrieren, wie genau das Kriterium zur Bestimmung der Faktorenanzahl im Bereich hoher Kommunalitäten arbeitet.



BEISPIEL 11

$N = 100$, $n = 15$, $k = 3$

Mehrfachladungen .80/.50/.00

$\bar{SMC} = .88$

$\bar{SMC}_{Limes} = .204$

Schätzung: $\bar{h}_S^2 = .885$

Gesamt- Bias

1 Faktor : $(1 - .885) 1/16 = .007 = 0,7\%$

2 Faktoren: $(1 - .885) 2/16 = .014 = 1,4\%$

3 Faktoren: $(1 - .885) 3/16 = .021 = 2,1\%$

4 Faktoren: $(1 - .885) 4/16 = .028 = 2,8\%$

Faktorenanzahl:

Mit dem dritten Faktor übersteigt die bias-bereinigte Varianz-Summe den geschätzten Wert.

2. Ein Anwendungsbeispiel aus der Literatur

Für eine Anwendung der neuen Technik der Faktorenanzahlbestimmung wurde Lienerts Untersuchung (G.A.Lienert, Belastung und Regression, Meisenheim 1964) herangezogen.

Auf der Basis der Differenzierungstheorie der Intelligenz (Garrett 1946 und Burt 1954) untersucht Lienert die Gültigkeit der Regressionshypothese in den Veränderungen der Faktorenstruktur des IST von Amthauer. Regression wird unter dem Einfluß von LSD erwartet. Entsprechend wurden 65 Studenten unter Noemalbedingungen und unter dem Einfluß von LSD mit dem IST getestet.

Lienert erhält (Zentroidmethode) eine 5-faktorielle Lösung unter Normalbedingungen, eine 4-faktorielle Lösung bei den Studenten unter LSD-Einfluß. Diese Lösungen kann er psychologisch sinnvoll im Hinblick auf die Regressionshypothese interpretieren.

An dieser Stelle sollen die Korrelationsmatrizen Lienerts noch einmal faktorenanalysiert werden. Zum Vergleich wird auch noch eine von Amthauer (1955) berichtet Korrelationsmatrix (N = 400) herangezogen.

Auf Seite 109 sind die jeweiligen Ergebnisse aus der TFA1 und die SMCs aufgeführt.

Auf Seite 110 sind die jeweiligen graphischen Lösungen zur Bestimmung der Faktorenanzahl nach dem entwickelten Verfahren zu finden.

Auf Seite 111 sind die nach der Faktorenanzahlentscheidung - je nur zwei Faktoren - ermittelten Ergebnisse nach einer Varianz-Rotation dargestellt.

Der $\bar{SMC}_{Limes}$ ist für alle drei Stichproben bedeutsam überschritten (.17 bei den Lienert- Stichproben, .08 bei der Amthauer- Stichprobe).

VARIABLE	EST COMMUNALITY (SHC)	FACTOR	EIGENVALUE	PCT OF VAR	CUM PCT
VAR001	.40559	1	3.18567	35.4	35.4
VAR002	.32215	2	1.44644	16.1	51.5
VAR003	.30906	3	.94051	10.5	61.9
VAR004	.26026	4	.80417	8.9	70.9
VAR005	.21959	5	.71589	8.0	78.8
VAR006	.54270	6	.64732	7.2	86.0
VAR007	.38329	7	.53518	5.9	91.9
VAR008	.31574	8	.44275	4.9	96.9
VAR009	.29491	9	.28207	3.1	100.0

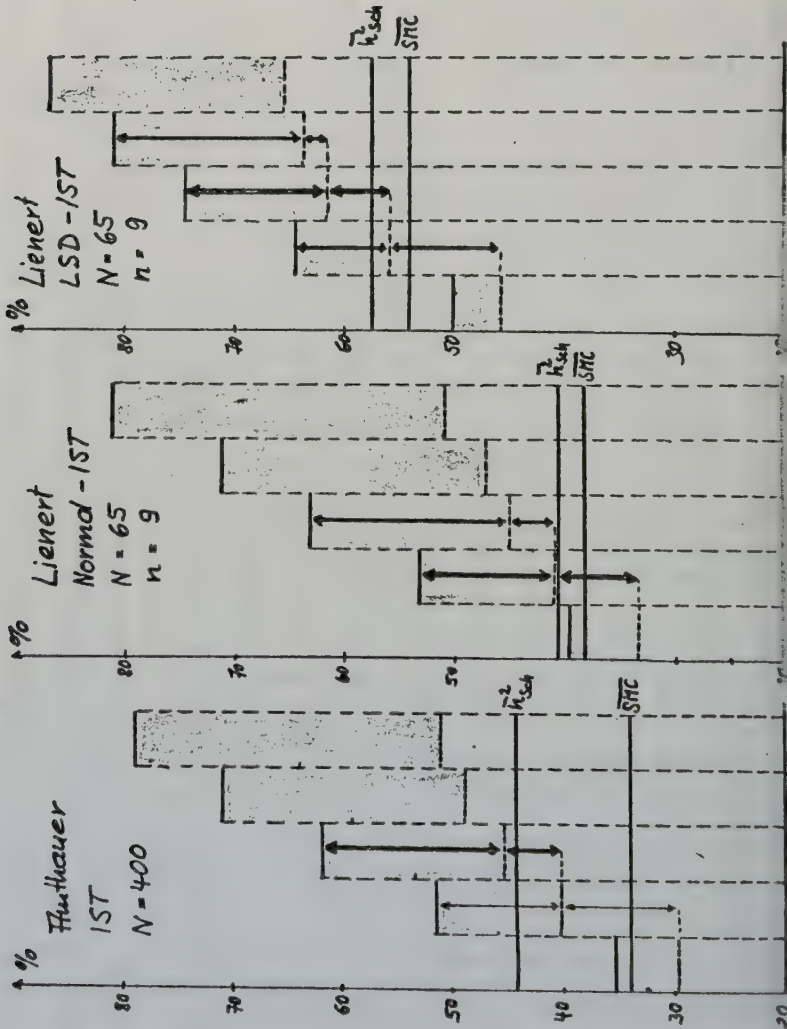
Lienert Normal

VARIABLE	EST COMMUNALITY (SHC)	FACTOR	EIGENVALUE	PCT OF VAR	CUM PCT
VAR001	.28072	1	3.53337	39.3	39.3
VAR002	.16523	2	1.23830	13.8	53.0
VAR003	.27326	3	.88814	9.9	62.9
VAR004	.15572	4	.83178	9.2	72.1
VAR005	.31847	5	.77978	8.7	80.8
VAR006	.70381	6	.63148	7.0	87.8
VAR007	.57544	7	.54444	6.0	93.9
VAR008	.53695	8	.36787	4.1	97.9
VAR009	.42243	9	.18525	2.1	100.0

Lienert LSD

VARIABLE	EST COMMUNALITY (SHC)	FACTOR	EIGENVALUE	PCT OF VAR	CUM PCT
VAR001	.59004	1	4.48804	49.9	49.9
VAR002	.39183	2	1.29895	14.4	64.3
VAR003	.69259	3	.90285	10.0	74.3
VAR004	.41914	4	.58057	6.5	80.8
VAR005	.43493	5	.52826	5.9	86.7
VAR006	.71545	6	.48958	5.4	92.1
VAR007	.59086	7	.31457	3.5	95.6
VAR008	.58312	8	.21248	2.4	98.0
VAR009	.43878	9	.18409	2.0	100.0

Anwendungsbeispiel 12 : Faktorenanzahlbestimmung zu Lienert ,
Regressionshypothese



Zu : Anwendungsbeispiel 12 : Rotierte Faktorladungsmatrizen

VARIMAX		ROTATED FACTOR MATRIX	
Amtheuer, IST , N = 400			
		FACTOR 1	FACTOR 2
VAR001	SE	.22955	.69029
VAR002	WA	.42719	.33545
VAR003	AN	.39836	.44884
VAR004	GE	.16703	.48341
VAR005	ME	.10513	.52235
VAR006	RA	.74075	.25043
VAR007	ZR	.41263	.33658
VAR008	FA	.63206	.06927
VAR009	WU	.63671	-.03586

VARIMAX		ROTATED FACTOR MATRIX	
Lienert, Normalbedingungen, N = 65			
		FACTOR 1	FACTOR 2
VAR001	SE	.18874	.51077
VAR002	WA	.07791	.45989
VAR003	AN	.32598	.48462
VAR004	GE	.08558	.43262
VAR005	ME	.39036	.34076
VAR006	RA	.76170	.41845
VAR007	ZR	.64462	.39670
VAR008	FA	.65573	.28078
VAR009	WU	.70976	-.02827

VARIMAX		ROTATED FACTOR MATRIX	
Lienert, LSD, N = 65			
		FACTOR 1	FACTOR 2
VAR001	SE	.68182	.24067
VAR002	WA	.65073	.14469
VAR003	AN	.81427	.33857
VAR004	GE	.66857	.17569
VAR005	ME	.44689	.31079
VAR006	RA	.59723	.61477
VAR007	ZR	.42479	.59333
VAR008	FA	.20780	.85277
VAR009	WU	.09930	.64531

Nach den aus den Simulationen gewonnenen Ergebnissen hat Lienert zu viele Faktoren extrahiert. Die Frage, warum seine Ergebnisse trotzdem sinnvoll interpretiert werden konnten, kann in einem ersten Ansatz folgendermaßen beantwortet werden : Zu dem eigentlichen Faktorenraum der zwei Faktoren wurde ein dreidimensionaler Raum größtenteils zufälliger gemeinsamer Varianz hinzugenommen. Die Rotation zur Einfachstruktur berücksichtigt dann die höchsten Interkorrelationen der Variablen. Das zeigt sich deutlich bei der 5-faktoriellen Lösung (Normalbedingungen). Vier Faktoren sind durch je zwei Variablen, einer nur durch eine (ME) Variable markiert.

Im vorliegenden Zusammenhang wird von Interesse, ob durch die geringere Faktorenanzahl noch Aussagen zur Regressionshypothese möglich sind. Da die rotierten Matrizen über die Hauptachsenanalyse (TFAK) gewonnen wurden, können sie als mit nur einem geringen Bias versehen betrachtet werden. Deshalb kann versucht werden, die beiden Faktoren jeweils nach ihren Varianzanteilen an der Gesamtvarianz zu beschreiben. Dadurch ist auch der Varianzzuwachs unter LSD berücksichtigt. Die Faktoren sind durch die Items des Tests (verbale gegenüber nicht-verbale Items) markiert. Dies soll bei der Varianzanteilbeschreibung berücksichtigt werden. Danach ergibt sich folgendes Bild :

	Normalbedingungen				LSD			
	Verbal	N-Verbal	Me	Summe	Verbal	N-Verbal	Me	Summe
1.Faktor	2 %	21 %	2 %	25 %	22 %	6 %	2 %	30 %
2.Faktor	10 %	4 %	1 %	15 %	2 %	21 %	1 %	24 %
Summe	12 %	25 %	3 %	40 %	24 %	27 %	3 %	54 %

Die Steigerung der gemeinsamen Varianz unter LSD ist demnach fast ausschließlich im verbalen Bereich zu finden. "Schließlich könnte man die Tatsache, daß das sprachliche Denken unter LSD so an Bedeutung gewinnt, mit der Tatsache in Zusammenhang bringen, daß der Spracherwerb als wichtigste Grundlage reifen menschlichen Denkens ontogenetisch früh die sensumotorische Intelligenz ablöst; unter Belastungsbedingungen greifen die Vpn offenbar so weit wie möglich auf diese Grundlage zurück." (Lienert 1964, S.60)

Weitergehende Aussagen sind auf der Basis der Daten Lienerts nicht möglich. Differenziertere Faktorenstrukturen können nur auf der Basis einer Steigerung der Variablenanzahl gewonnen werden.

3. Ein Anwendungsbeispiel aus der Literatur

Die Faktorenstruktur des FPI

Grundlage der folgenden Betrachtungen sind die in der Handanweisung des Freiburger Persönlichkeitsinventars (Fahrenberg und Selg, 1970) berichteten Eigenwertreihen.

Da von den Autoren keine SMC-Werte berichtet werden, muß sich eine Faktorenanzahlbestimmung an der Bias-Berechnung für die TFA1 orientieren. Bisher wurde eine Doppelstrategie benutzt :

1. Schätzung von $\bar{h}_S^2$ über die $\bar{SMC}$
2. Bias - Bereinigung der TFA1 - Varianzsummen aus den Eigenwerten

Ohne die Möglichkeit, aus dem $\bar{SMC}$ -Wert die mittlere Kommunalität zu schätzen, muß eine Technik der schrittweisen Näherung an die 'wahre' Faktorenanzahl eingeschlagen werden. Dabei wird das über eine andere Technik der Faktorenanzahlbestimmung gewonnen $\bar{h}^2$ zu einer ersten Bias - Bereinigung der TFA1 - Ergebnisse benutzt. Über die Beurteilung von Zuwachs an bias-bereinigter Varianz und Gesamt-Bias kann ein neues $\bar{h}^2$ geschätzt werden. Auf dessen Basis muß wieder eine Bias-Bereinigung durchgeführt werden.

A. Erste Analysenstichprobe des FPI, N = 630 , n = 240

(Darstellung Seite 114)

Mittlere Kommunalität der 9-faktoriellen Lösung : $\bar{h}_{TFA1}^2 = .26$

Die nach diesem $\bar{h}^2$ bias-bereinigte Summenkurve (% der Spur)

beschreibt erst mit dem 13.Faktor 26% der Gesamtvarianz.

Eine $\bar{h}^2$ -Schätzung nach der bias-bereinigten Kurve läßt bei dem

6.Faktor den Zuwachs im Vergleich zu dem Gesamtbias (Pfeile)

gerade noch groß genug erscheinen. Nach einer neuerlichen Biasbestimmung muß erwartet werden, daß sich dieses Verhältnis ver-

ändert, und zwar zu ungunsten eines Varianzzuwachses des 6.Fak-

tors. Deshalb wird das neue $\bar{h}^2$ als $\bar{h}^2 = .19$ geschätzt. Nach der

danach neu durchgeführten Bias-Bereinigung zeigt sich, daß der

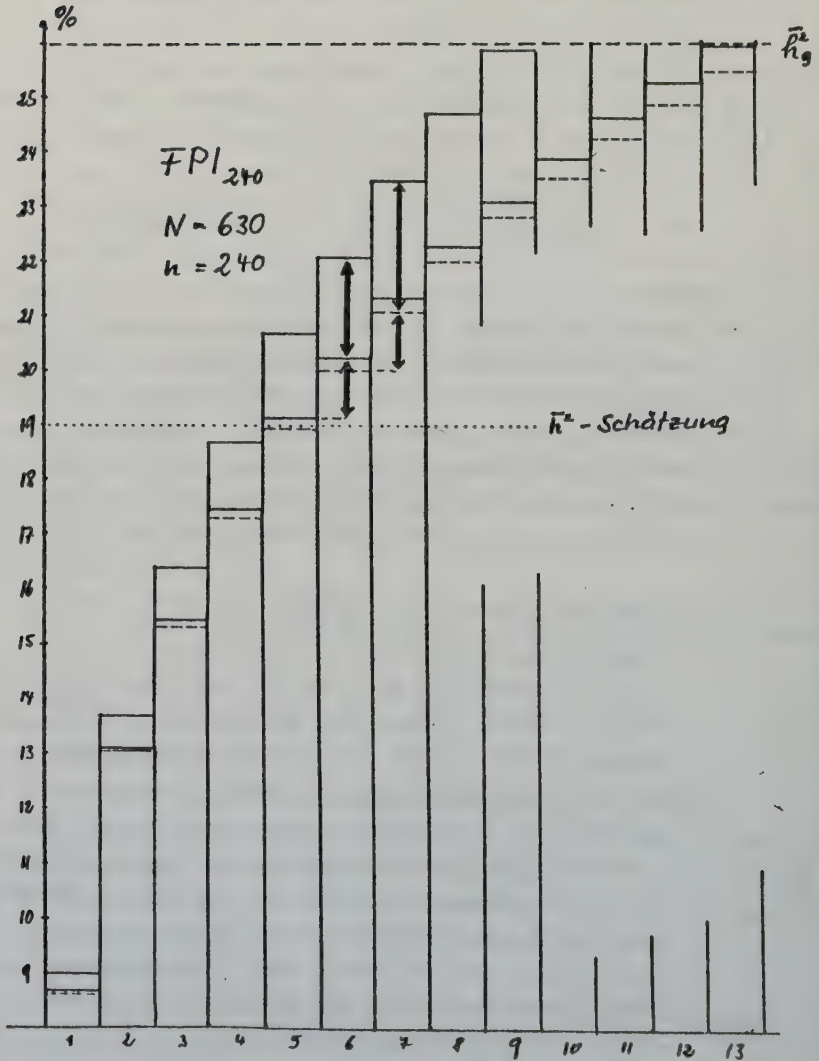
6.Faktor noch als echter Faktor anzusprechen ist. Beim 7.Faktor dagegen ist der Zuwachs im Verhältnis zum Bias zu gering (Pfeile).

Das Stichproben- $\bar{h}^2$ muß demnach zwischen .19 und .20 liegen.

Bestimmung der Faktorenanzahl

FPI, 1. Analysenstichprobe,

$N = 630$, $n = 240$



B. Faktorenanalyse der FPI-Standardskalen (N = 630 , n = 9)

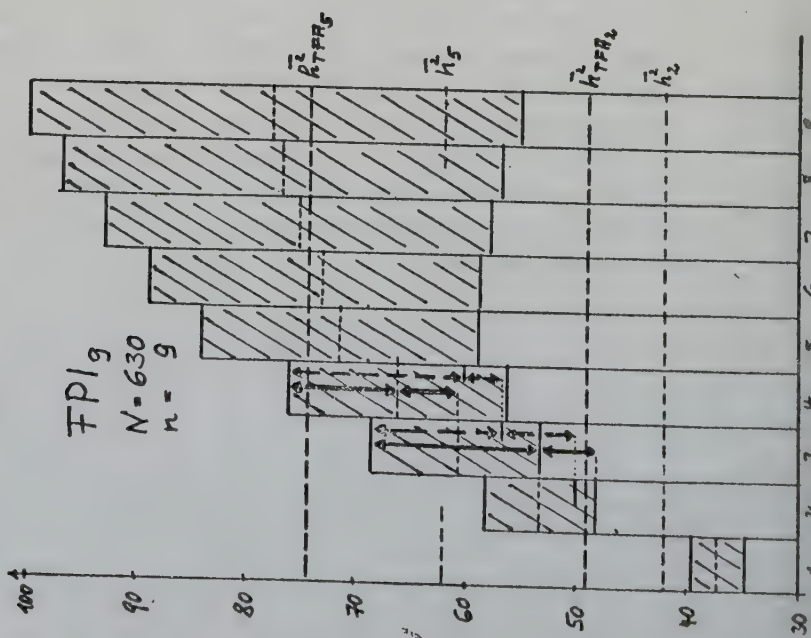
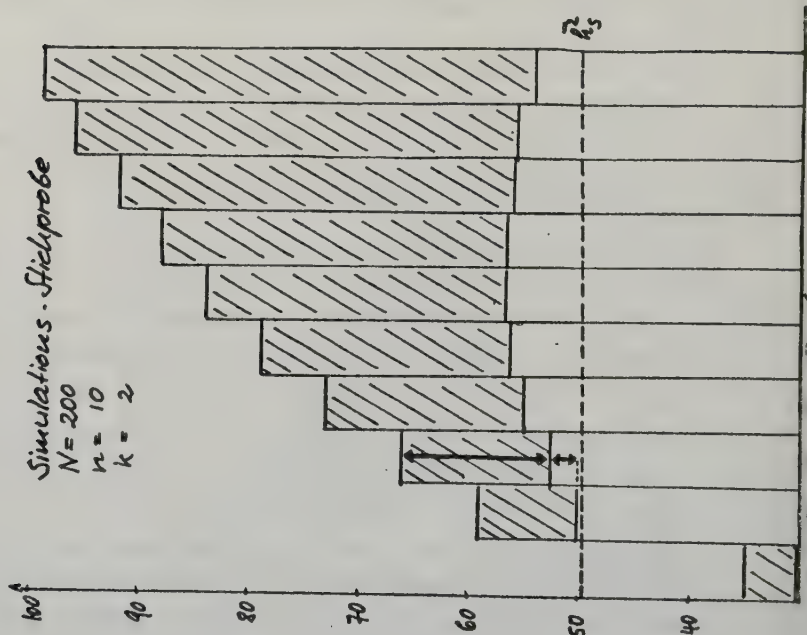
Die Autoren haben sich hier für eine 5-Faktoren-Lösung als die psychologisch fruchtbarere entschieden. Außerdem wird begründet, daß ein größerer Varianzanteil aufgeklärt würde, wie es angesichts der Reliabilitätsschätzungen auch nur sinnvoll sei (S.39 a.a.O.). Eine 2-Faktoren-Lösung wird zusätzlich berichtet.

Ausgehend von der 5-Faktoren-Lösung und dessen $\bar{h}_{TFA_5}^2$ läßt sich nach der Bias-Bereinigung ein neues $\bar{h}_5^2 = .61$ bestimmen, das den Zuwachs an Varianz im Verhältnis zum Bias berücksichtigt. Wird danach wieder der Bias bestimmt, erhält man eine Faktorenanzahl von maximal 3 Faktoren (s. unterbrochene Pfeile im 3. und 4.Feld, S.116).

Führt man eine Bias-Bereinigung über die mittlere Kommunalität der 2-Faktoren-Lösung durch, dann erscheint der dritte Faktor fraglich (durchgehende Pfeile im dritten Feld).

Der Darstellung der obigen graphischen Lösungsbeschreibung ist eine vergleichbare Stichprobe aus den simulierten Daten gegenübergestellt (Seite 116). Auch wenn hier die Entscheidung für zwei der drei Faktoren nicht endgültig gefällt werden kann, solange keine Information über $\bar{S}\bar{M}\bar{C}$ zur Schätzung der mittleren Kommunalität in der Stichprobe vorliegt, ist die Ablehnung einer 5-Faktoren-Lösung eindeutig. Eine Begründung dieser Lösung über die Reliabilitätsschätzungen kann nicht akzeptiert werden. Für diese Reliabilitäten sind nicht nur die Anteile der Variablen am Raum der gemeinsamen Varianz, sondern auch deren spezifische Varianzanteile bedeutsam. Erst für h^2 gegen 1 kann erwartet werden, daß Kommunalität und Reliabilität übereinstimmen.

Bestimmung der Faktorenenzahl: FPI-Standardskalen



VII. Zusammenfassung

Auf der Basis der Erkenntnisse, daß die Faktorenanalyse den Ansprüchen genügt (Oberla 1968 ; Schiller u. Wittmann 1971), wurde das Ziel verfolgt, den systematischen Fehler bei einigen faktorenanalytischen Verfahren zu beschreiben.

Gegenüber anderen Modellrechnungen (Box-Problem Thurstone; Ball-Problem Cattell u. Dickman) sollte die mögliche Generalisierbarkeit bei diesen Monte-Carlo-Experimenten genutzt werden. Es handelt sich um Stichprobenexperimente aus definierten Grundgesamtheiten.

Zwei Arten systematischer Fehler wurden gefunden, beschrieben und abgeleitet.

Für die Hauptkomponentenanalyse und die Multiplen Korrelationen wurde der erste systematische Fehler formelmäßig erfaßt.

Die Konsequenz ist ein neues Verfahren zur Bestimmung der Anzahl der Faktoren. Dabei wird aus der Möglichkeit zur Bestimmung des systematischen Fehlers Nutzen gezogen, und die mittlere gemeinsame Varianz in der Stichprobe in Abhängigkeit von den Stichprobenkennwerten bestimmt. Diese Varianz entspricht der durch die Faktoren zu extrahierenden Varianz.

Für die Praxis liegt die Bedeutung des neuen Verfahrens darin, daß ein 'blindes' Bestimmen der Faktorenanzahl, zum Beispiel ohne Berücksichtigung der Interpretierbarkeit der Faktoren, ermöglicht wird.

Die Daten

Der größte Teil der Simulationen und Faktorenanalysen wurde am Rechenzentrum der Universität Freiburg durchgeführt. Einzelne ergänzende Berechnungen wurden am Rechenzentrum der Universität Karlsruhe durchgeführt. Dabei diente das 'SPSS' (Statistical Package for the Social Sciences, Chi Corporation Ohio) als Programmband.

Eine Dokumentation der gesamten Daten ist bei dem Umfang der simulierten Stichproben nicht durchführbar. Deshalb werden nur die Mittelwerte der Kommunalitäten, getrennt nach der Grundlage (Eine bzw. fünf Stichproben) in den 'Daten 1' berichtet. 'Daten 2' sind die auf vergleichbare Basis bei den Stichproben-Kommunalitäten umgerechnete Werte. 'Daten 3' sind die jeweiligen Abweichungen der mittleren Kommunalitäten in den Ergebnissen von den mittleren Stichprobenkommunalitäten auf der Grundlage von 'Daten 2'.

Ergänzend werden sieben Auswertungsbeispiele dargestellt.

Daten 1

N/n/k Stichpr.	25/10/2 5	50/10/2 1	100/10/2 1	5	1	200/10/2 5	1	200/7+3/2 5	1	200/6/2 5	1	100/15/3 1
$\bar{h}_S^2$				824		833						860
TFA1				858		850						914
TFAK				823		812						892
SMC				796		777						878
IFA				779		763						865
AFA				823		812						892
KFA				800		798						863
$\bar{h}_S^2$				664		668						588
TFA1				732		715						697
TFAK				666		645						616
SMC				619		584						575
IFA				589		560						531
AFA				666		645						616
KFA				666		642						610
$\bar{h}_S^2$	520	463	507	489	515	486	500	484	481	495	544	481
TFA1	618	600	585	594	622	586	589	582	584	666	680	619
TFAK			504		544		487		480			525
SMC	601	560	459	436	470	403	472	400	396	334	365	479
IFA	472	437	404	395	429	375	379	366	366	290	324	431
AFA		506	502		528		487		473		521	525
KFA		506	469		540		479		482		422	520
$\bar{h}_S^2$				387		392						313
TFA				527		493						483
TFAK				410		369						358
SMC				347		288						312
IFA				301		252						251
AFA				410		371						358
KFA				395		335						338
$\bar{h}_S^2$	299	289		266	275	250	263	243	247	258	299	257
TFA1	453	448		414	448	388	397	388	398	500	522	385
TFAK		319			312				251			287
SMC	466	412		231	252	167	180	169	180	119	144	268
IFA	287	252		172	202	132	140	133	150	194	113	205
AFA		322			313		257		248		298	206
KFA		302			282		213		233		196	204
$\bar{h}_S^2$	217	204		179	186	162	174	161		169	203	
TFA1	396	336		340	385	310	317	323		440	462	
TFAK		250					176					
SMC	426	355		165	183	101	113	104		67	82	
IFA	227	181		99	131	69	74	71		37	58	
AFA		327			232		184				215	
KFA		283			235		217				194	
$\bar{h}_S^2$	152			109		95	104	90		99		
TFA1	363			292		268	281	278		420		
SMC	404			125		63	75	70		40		
IFA	203			62		36	45	39		29		

n=14, k=3

n=14, k=3

Daten 2

N/n/k Stichpr.	25/10/2 5	50/10/2 1	100/10/2 1	200/10/2 5	200/10/2 1	200/7+3/2 5	200/7+3/2 1	200/6/2 5	200/6/2 1	100/15 1
$\bar{h}_S^2$				810	810					810
TFA1				843	827					861
TFAK				809	790					840
SMC				782	756					827
IFA				766	742					815
AFA				809	790					840
KFA				794	772					813
$\bar{h}_S^2$				640	640					640
TFA1				705	685					759
TFAK				642	618					670
SMC				597	560					626
IFA				568	537					578
AFA				642	618					670
KFA				642	615					664
$\bar{h}_S^2$	490	490	490	490	490	490	490	490	490	490
TFA1	582	634	565	594	592	591	577	589	591	630
TFAK			487	522	522		477		491	535
SMC	566	592	444	436	447	406	404	405	403	488
IFA	445	462	390	395	408	378	371	370	373	439
AFA		536	485	502	477		477		482	535
KFA		536	453	521	469		469		491	530
$\bar{h}_S^2$				360	360					360
TFA1				490	453					556
TFAK				381	339					412
SMC				323	264					359
IFA				280	231					289
AFA				381	341					412
KFA				367	308					389
$\bar{h}_S^2$	250	250		250	250	250	250	250	250	250
TFA1	379	387		389	407	388	377	388	403	375
TFAK		275		284	284				254	279
SMC	390	356		217	229	167	171	169	182	261
IFA	240	218		162	184	132	133	133	152	229
AFA		278		285	285		244		251	278
KFA		261		256	256		202		236	280
$\bar{h}_S^2$	160	160		160	160	160	160	160	160	160
TFA1	292	264		304	331	308	291	323	416	364
TFAK		196					162			
SMC	314	278		147	157	101	104	104	63	65
IFA	167	142		88	113	69	68	71	35	46
AFA		278		204	204		169			169
KFA		256		203	203		117			153
$\bar{h}_S^2$	90			90		90	90	90		90
TFA1	215			241		253	243	278	382	297
SMC	239			103		60	65	70	36	32
IFA	120			51		35	39	39	26	22

Daten 3

N/n/k Stichpr.	25/10/2 5	50/10/2 1	100/10/2 5	200/10/2 1	200/7+3/2 5	200/6/2 1	100/15/3 1
$\bar{r}_S^2$			81o	81o			81o
TFA1			+ 33	+ 17			+ 51
TFAK			- 1	- 2o			+ 3o
SMC			- 28	- 54			+ 17
IFA			- 44	- 68			+ 5
AFA			- 1	- 2o			+ 3o
KFA			- 16	- 38			+ 3
$\bar{r}_S^2$			64o	64o			64o
TFA1			+ 85	+ 45			+119
TFAK			+ 2	- 22			+ 3o
SMC			- 43	- 8o			- 14
IFA			- 72	-1o3			- 62
AFA			+ 2	- 22			+ 3o
KFA			+ 2	- 25			+ 2o
$\bar{r}_S^2$	49o 49o	49o	49o 49o	49o 49o	49o 49o	49o 49o	49o
TFA1	+ 92 +144	+ 75	+1o4 +1o2	+1o1 + 87	+ 99 +1o1	+169 +122	+14o
TFAK		- 3	+ 32	- 12	+ 1		+ 45
SMC	+ 76 +1o2	- 46	- 54 - 43	- 84 - 86	- 85 - 87	-16o -161	- 1
IFA	- 45 - 28	-1oo	- 95 - 82	-112 -119	-12o -117	-2o3 -198	- 51
AFA	+ 46 - 5	- 5	+ 12	- 13	- 8	- 21	+ 45
KFA	+ 46	- 37	+ 31	- 21	+ 1	-11o	+ 4o
$\bar{r}_S^2$			36o	36o			36o
TFA1			+13o	+ 97			+196
TFAK			+ 21	- 21			+ 52
SMC			- 37	- 96			- 1
IFA			- 8o	-129			- 71
AFA			+ 21	- 19			+ 52
KFA			+ 7	- 52			+ 29
$\bar{r}_S^2$	25o 25o		25o 25o	25o 25o	25o 25o	25o 25o	25o
TFA1	+129 +137		+139 +157	+133 +127	+138 +153	+234 +186	+125
TFAK	+ 25		+ 34		+ 4		+ 29
SMC	+14o +1o6		- 33 - 21	- 83 - 79	- 81 - 68	-135 -13o	+ 11
IFA	- 1o - 32		- 88 - 66	-118 -117	-117 - 98	-159 -156	- 21
AFA	+ 28		+ 35	- 6	+ 1	- 1	+ 28
KFA	+ 11		+ 6	- 48	- 14	+ 14	+ 26
$\bar{r}_S^2$	16o 16o		16o 16o	16o 16o	16o	16o 16o	
TFA1	+132 +1o4		+144 +177	+149 +131	+163	+256 +2o4	
TFAK	+ 36			+ 2			
SMC	+154 +118		- 13 - 3	- 59 - 56	- 56	- 97 - 95	
IFA	+ 7 - 18		- 72 - 47	- 91 - 92	- 89	-125 -114	
AFA	+118		+ 44	+ 9		+ 9	
KFA	+ 96		+ 43	- 43		- 7	
$\bar{r}_S^2$	9o		9o	9o 9o	9o	9o 9o	
TFA1	+125		+151	+163 +153	+168	+272 +2o7	
SMC	+149		+ 13	- 3o - 25	- 2o	- 54 - 58	
IFA	+ 3o		- 39	- 55 - 51	- 51	- 64 - 68	

k=3

k=2

AUSWERTUNGSBEISPIEL 1

N = 100 , n = 10 , k = 2 , 4 Markiervariable: Var.Nr. 1,2,6,7

A1 Kommunalitäten

Var.Nr.	1	2	3	4	5	6	7	8	9	10	Σh_i^2
h_S^2	451	598	332	512	445	657	654	504	552	396	510
h_{TFA1}^2	597	642	585	646	603	732	733	641	638	520	634
h_{TFAK}^2	484	546	501	570	508	682	682	529	534	402	544
SMC	444	466	529	505	451	582	593	479	510	380	490
h_{IFA}^2	393	431	442	483	411	552	554	431	454	356	451
h_{AFA}^2	482	549	486	571	519	687	685	529	528	402	544
h_{KFA}^2	486	507	545	577	524	611	610	528	518	516	544

B. Nivellierung

Var.Nr.	1	2	3	4	5	6	7	8	9	10	h_i^2
$h_S^2 - h_S^2$	59	88	-178	2	-64	147	144	-6	42	-114	120
TFA1	22	-80	129	10	33	-49	-45	13	-38	0	120
TFAK	1	-86	135	24	28	-9	-6	-9	-52	-28	30
SMC	9	-116	213	9	21	-59	-45	-9	-26	0	10
IFA	1	-108	169	30	24	-46	-41	-14	-39	19	5
AFA	3	-83	120	25	39	-4	-3	-9	-58	-28	30
KFA	3	-123	181	33	46	-78	-76	-8	-66	88	32

AUSWERTUNGSBEISPIEL 2

$N = 100$, $n = 10$, $k = 2$, Einfachladungen (Population) .70

A. Kommunalitäten

Var.Nr.	1	2	3	4	5	6	7	8	9	10	$\sum h^2/n$
h_s^2	480	378	428	322	735	410	381	582	494	584	479
h_{TFA1}^2	569	561	575	517	673	563	576	586	569	632	582
h_{TFAK}^2	507	444	429	376	619	466	416	521	480	542	480
SMC	427	378	385	334	481	389	371	414	429	445	405
h_{IFA}^2	402	354	333	310	443	344	330	383	387	408	372
h_{AFA}^2	504	457	421	376	618	470	412	522	474	547	480
h_{KFA}^2	491	476	477	463	509	481	471	492	485	494	484
MSA	922	920	920	922	925	929	914	939	908	920	922

B. Nivellierung

Var.Nr.	1	2	3	4	5	6	7	8	9	10	$\sum h^2/n - h_s^2$
$h_s^2 - h_s^2$	1	-101	-51	-157	256	-69	-98	103	15	105	
TFA1	14	80	44	92	-165	50	92	-99	-28	-55	103
TFAK	26	65	0	54	-117	55	34	-62	-15	-43	1
SMC	21	74	31	86	-180	53	64	-94	9	-65	-74
AFA	23	78	-8	53	-118	59	30	-61	-21	-38	1
KFA	6	93	44	136	-231	66	85	-95	-14	-95	5
MSA	1	99	49	157	-253	76	90	-86	-29	-107	443

AUSWERTUNGSBEISPIEL 3

N = 200 , n = 10 , k = 2 , Einfachladungen (Population) .70

1.Faktor 7 Variablen

2.Faktor 3 Variablen

A. Kommunalitäten

Var.Nr.	1	2	3	4	5	6	7	8	9	10	h ²
h _S ²	559	512	514	480	540	582	426	297	483	420	4
h _{TFA1} ²	626	591	552	574	626	552	560	541	605	608	1
SMC	492	471	427	454	494	436	450	236	234	264	
h _{IFA} ²	473	446	411	433	468	412	419	196	189	211	
h _{AFA} ²	566	526	477	504	561	477	482	333	386	431	
h _{KFA} ²	578	566	550	560	573	550	553	289	273	330	
MSA	965	969	973	969	963	968	962	749	651	739	

B. Nivellierung

Var.Nr.	1	2	3	4	5	6	7	8	9	10	h ²
h _S ² - h _S ⁻²	78	31	33	- 1	59	101	- 55	-184	2	- 61	
TFA1	- 36	- 24	- 65	- 9	- 17	-133	31	141	19	85	
SMC	18	44	- 2	59	39	- 61	109	24	-164	- 71	
IFA	- 26	6	21	47	- 21	- 65	78	108	-101	- 48	
AFA	15	22	- 23	32	21	- 97	56	44	- 89	19	
KFA	18	53	36	80	79	33	127	- 8	-210	- 90	
MSA	- 4	47	49	79	13	- 24	126	42	-242	- 91	

AUSWEICHUNGSBEISPIEL 4

N = 200 , n = 10 , k = 2 , Einfachladungen (Population) .50

1.Faktor 7 Variablen

2.Faktor 3 Variablen

A. Kommunalitäten

Var.Nr.	1	2	3	4	5	6	7	8	9	10	$\Sigma h^2/n$
h_S^2	312	270	271	244	295	337	199	98	248	191	247
h_{TFA1}^2	492	463	364	377	490	380	453	147	326	490	398
h_{TFAK}^2	379	366	250	270	358	268	299	58	53	210	251
SMC	256	244	191	220	262	195	221	78	45	85	180
h_{IFA}^2	228	210	171	198	223	177	177	39	34	47	150
h_{AFA}^2	371	328	264	275	382	235	292	71	78	185	248
h_{KFA}^2	330	320	272	288	315	274	287	71	37	133	233
MSA	866	849	858	845	816	845	819	402	73	568	654

B. Nivellierung

Var.Nr.	1	2	3	4	5	6	7	8	9	10	$\bar{h}_A^2 - \bar{h}_S^2$
$h_S^2 - \bar{h}_S^2$	65	23	24	- 3	48	90	- 48	-149	1	- 56	
TFA1	29	42	- 58	- 18	44	-108	103	-102	- 73	148	151
TFAK	63	92	- 25	22	59	- 73	96	- 44	-199	15	4
SMC	11	41	- 13	43	34	- 75	89	47	-136	- 39	- 67
h_{IFA}^2	14	38	- 2	52	26	- 62	76	39	-116	- 46	- 98
h_{AFA}^2	58	57	- 8	30	86	-102	93	- 28	-171	- 7	1
h_{KFA}^2	32	64	15	58	34	- 49	102	- 13	-197	- 44	- 13
MSA	107	132	140	154	74	64	173	-143	-622	- 70	447

AUSWERTUNGSBEISPIEL 5

N = 200 , n = 10 , k = 2 , Einfachladungen (Population) .80

A. Kommunalitäten

Var.Nr.	1	2	3	4	5	6	7	8	9	10	Σ
h_S^2	705	661	661	629	688	528	572	870	654	715	6
h_{TFA1}^2	750	721	712	707	733	695	665	765	672	734	7
h_{TFAK}^2	695	651	638	632	669	613	575	719	583	670	6
SMC.	622	593	585	571	608	556	555	639	519	597	5
h_{IFA}^2	598	567	561	554	581	536	521	603	509	570	4
h_{AFA}^2	696	653	638	631	669	610	571	723	585	672	5
h_{KFA}^2	662	650	649	648	655	627	620	651	621	640	4

B. Nivellierung

Var.Nr.	1	2	3	4	5	6	7	8	9	10	h
$h_S^2 - h_S^2$	37	- 7	- 7	- 39	20	-140	- 96	202	- 14	47	-
TFA1	2	13	4	31	- 2	120	46	-154	- 29	- 28	-
TFAK	11	13	0	26	4	108	26	-128	- 48	- 22	-
SMC.	1	16	8	26	4	112	67	-147	- 51	- 34	-
IFA	1	14	8	33	1	116	57	-159	- 37	- 37	-
AFA	14	15	0	25	4	105	22	-124	- 46	- 20	-
KFA	17	15	14	45	- 7	125	74	-193	- 7	- 49	-

AUSWERTUNGSBEISPIEL 6

N = 25 , n = 10 , k = 2 , Populationsfaktorladungen (einfach) .50

A. Kommunalitäten

Var.Nr.	1	2	3	4	5	6	7	8	9	10	$\sum h^2/n$
h_S^2	329	279	166	254	316	248	317	350	467	161	289
h_{TFA1}^2	580	246	477	368	579	494	326	376	588	445	448
h_{TFAK}^2	550	199	260	186	420	305	216	228	527	297	319
SMC	592	382	429	377	446	315	468	312	508	290	412
h_{IFA}^2	387	194	217	179	334	183	245	216	325	235	252
h_{AFA}^2	469	181	241	236	450	317	175	275	610	262	322
h_{KFA}^2	494	284	254	163	386	226	310	203	429	273	302
MSA	759	532	530	434	776	638	492	512	702	629	625

B. Nivellierung

Var.Nr.	1	2	3	4	5	6	7	8	9	10	$\sum h_A^2 - h_S^2$
$h_S^2 - h_S^2$	40	- 10	-123	- 35	27	- 41	28	61	178	-128	
TFA1	92	-192	152	- 45	104	87	-150	-133	- 38	125	159
TFAK	191	-110	64	- 98	74	27	-131	-152	30	106	30
SMC	140	- 20	140	0	7	- 56	28	-161	- 82	6	123
IFA	95	- 48	88	- 38	55	- 28	- 35	- 97	-105	111	- 37
AFA	107	-131	42	- 51	101	36	-175	-108	110	68	33
KFA	152	- 8	75	-104	57	- 35	- 20	-160	- 51	99	13
MSA	94	- 83	28	-156	124	54	-161	-174	-101	132	336

AUSWERUNGSBEISPIEL 7

$N = 100$, $n = 15$, $k = 3$, Mehrfachladungen (Population) .50 .30 .00

A. Kommunalitäten

Var.Nr.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	h^2/n
h_S^2	265	315	235	270	260	251	470	337	271	460	322	285	232	353	366	313
h_{TFA1}^2	427	634	404	455	497	545	562	492	502	381	448	450	541	417	489	483
h_{TFAK}^2	288	593	236	308	314	427	492	377	367	265	279	333	463	253	378	358
S.I.C.	264	445	191	268	264	332	428	379	282	280	248	261	390	286	366	312
h_{TFA}^2	229	334	169	213	226	287	354	298	256	214	208	234	299	174	278	251
h_{AFA}^2	280	572	242	327	327	411	499	361	388	262	251	341	453	266	388	358
h_{KFA}^2	397	478	380	394	394	430	466	436	428	395	389	414	441	362	436	416
MSA	806	801	848	793	783	813	858	830	876	801	792	873	769	709	805	814

B. Nivellierung (Fortsetzung Auswertungsbeispiel 7)

Var. Nr.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	$\bar{h}_A^2 - \bar{h}_S^2$
$\bar{h}_S^2 - \bar{h}_S^2$	- 48	2 - 78	- 43	- 53	- 62	157	24 - 42	147	9 - 28	- 81	40	53				
TFA1	- 8	149 - 1	15	67	124	- 78	- 15	61 - 249	- 44 - 5	139	- 106	- 47	170			
TFAK	- 22	233 - 44	- 7	9	131	- 23	- 5	51 - 240	- 88	3	186	- 145	- 33	45		
SAC	3	134 - 40	2	8	85	- 38	46	15 - 176	- 70 - 20	162	- 63	4	- 4			
IFA	23	81 - 4	5	28	98	- 54	23	47 - 184	- 52	11	129	- 117	- 26	- 62		
AFA	- 30	212 - 38	12	22	115	- 16	- 21	72 - 243	- 116	11	167	- 132	- 23	45		
KFA	29	60	42	21	31	76	- 107	- 4	54 - 168	- 36	26	106	- 94	- 33	103	
MSA	40	- 15	113	23	23	62	- 112	- 7	105 - 160	- 30	88	36	- 144	- 60	501	

Beispiele für Populationsfaktorladungsmatrizen

A. Einfachladungen; $n = 10$, $k = 2$;

	1	2
1	.600	.000
2	-.600	.000
3	.600	.000
4	-.600	.000
5	.600	.000
6	.000	-.600
7	.000	.600
8	.000	-.600
9	.000	.600
10	.000	-.600

B. Mehrfachladungen mit Markiertvariablen; $n = 10$, $k = 2$

	1	2
1	0.700	-0.000
2	-0.700	-0.000
3	0.500	-0.500
4	-0.400	0.600
5	0.600	-0.400
6	-0.000	0.800
7	-0.000	-0.800
8	0.500	0.500
9	-0.600	-0.400
10	0.400	-0.600

C. Mehrfachladungen .60 .40 .00 ; $n = 15$, $k = 3$

	1	2	3
1	0.600	0.400	-0.000
2	-0.600	0.400	-0.000
3	0.600	-0.400	-0.000
4	-0.600	-0.000	-0.400
5	0.600	-0.000	-0.400
6	-0.000	0.600	0.400
7	-0.000	-0.600	0.400
8	-0.000	0.600	-0.400
9	0.400	-0.600	-0.000
10	0.400	0.600	-0.000
11	-0.400	-0.000	0.600
12	-0.400	-0.000	-0.600
13	0.400	-0.000	0.600
14	-0.000	-0.400	-0.600
15	-0.000	-0.400	0.600

LITERATUR

- Anderson, T.W. (1958): Introduction to multivariate statistical analysis, NY: Wiley
- Bentler, P.M. (1968): Alpha-maximized factor analysis (alphamax): Its relation to alpha and canonical factor analysis. Psychom. 33
- Bentler, P.M. (1969): Some extensions of Image Analysis. Psychom. 34
- Browne, M.W. (1968): A comparison of factor analytic techniques. Psychom. 33
- Browne, M.W. (1969): Fitting the factor analysis model. Psychom. 34
- Cattell, R.B. & Dickman (1962a): A dynamic model of physical influences demonstrating the necessity of oblique simple structure. Psychol. Bull. 59
- Cattell, R.B. (1962b): The basis of recognition and interpretation of factors. Ed. and Psych. Measurement, 22
- Cattell, R.B. (1966a): Handbook of multivariate experimental psychology. Chicago
- Cattell, R.B. (1966b): The scree-test for the number of factors. Multiv. Beh. Res. 1
- Cliff, N. & Pennell, R. (1967): The influence of communality, factor strength, and loading size on the sampling characteristics of factor loadings. Psychom. 32
- Cooley, W.W. & Lohnes, P.R. (1971): Multivariate data analysis. New York
- Darroch, J.N. (1965): A set of inequalities in factor analysis. Psychom. 30
- Edwards, A.L. (1971): Versuchsplanung in der psychologischen Forschung. Weinheim
- Fahrenberg, J. & Selg, H. (1970): Das Freiburger Persönlichkeitsinventar. Göttingen
- Fischer, G. & Roppert, J. (1965) Monte-Carlo-Untersuchungen an einem Faktorenmodell. Schriftenreihe des Instituts für höhere Studien und wissenschaftliche Forschung. Wien, Band 1
- Fischer, G.H. (1968): Psychologische Testtheorie. Bern
- Fischer, G.H. (1974): Einführung in die Theorie psychologischer Tests. Bern
- Fisher, R.A. (1950): Contributions to mathematical statistics. New York
Paper 12: On an distribution yielding the error functions of several well known statistics.
- Gaensslen, H. & Schubö, W. (1973): Einfache und komplexe statistische Analyse. München, Basel
- Gorsuch, R.L. (1974): Factor analysis. Philadelphia
- Guilford, J.P. (1965): Fundamental statistics in psychology and education. New York, 4th.ed.
- Gullahorn, J.E. (1967): Multivariate approaches in survey data processing: Comparison of factor, cluster, and Guttman analyses and of multiple regression and canonical correlation methods. Multivariate Behavioral Res. Mon.

- Guttman,L.(1956): Best possible systematic estimates of communalities. Psychom.21
- Guttman,L.(1953): Image theory for the structure of quantitative variates. Psychom.4
- Guttman,L.(1957): Simple proofs of relations between the communality problem and multiple correlation.Psychom.22
- Harman,H.(1967): Modern factor analysis.Chicago, 2nd.ed.
- Harris,C.W.(1962): Some Rao-Guttman relationships.Psychom.27
- Horn,J.L.(1965): A rationale and test for the number of factors in factor analysis.Psychom.30
- Horst,P.(1965): Factor analysis of data matrices. New York
- Humphreys,L.G.(1964): Number of cases and number of factors: an example where N is very large.Ed.and Psych.Measurement, 24
- Jäger,A.O.(1967): Dimensionen der Intelligenz. Göttingen
- Jöreskog,K.G.(1969): Efficient estimation in image factor analysis. Psychom.34
- Kaiser,H.F.(1961): A note on Guttman lower bound for the number of common factors. Ed.and Psychol.Measurement, 19
- Kaiser,H.F.(1963): Image analysis. In: Harris,C.W.: Problems in measuring change. Madison:University of Wisconsin Press
- Kaiser,H.F.(1970): A second generation little jiffy.Psychom.35
- Kaiser,H.F.& Caffrey,J.(1965): Alpha factor analysis.Psychom.30
- Kaiser,H.F.& Rice,J.(1974): Little Jiffy mark IV..Ed.and Psychol. measurement, 34
- Kalveram,K.T.(1965): Veränderung der Faktorenstruktur durch simultane Überlagerung. Arch.ges.Psychol.117
- Kalveram,K.T.(1970): Über Faktorenanalyse. Kritik eines theoretischen Konzepts und seine theoretische Neuformulierung. Archiv für Psychologie,122
- Lienert,G.A.(1964): Belastung und Regression. Meisenheim
- Lienert,G.A.(1967): Testaufbau und Testanalyse. Weinheim, 2.Aufl.
- Linn,R.L.(1968): a monte carlo approach to the number of factors problem. Psychom.33
- McLeod,J.(Ed.)(1968): Simulation. New York
- Menges,G.(1968): Statistik I, Theorie.Köln
- Naylor,T.H. et al.(1966): Computer simulation techniques. New York
- Novick,M.R.& Lewiss,C.(1967): Coefficient Alpha and the reliability of composite measurements.Psychom.32
- Orlik,P.(1967): Das Dilemma der Faktorenanalyse - Zeichen einer Aufbaukrise in der modernen Psychologie.
- Pawlik,K.(1968): Dimensionen des Verhaltens. Bern/Stuttgart

- Penell,R.(1968): The influence of communality and N on the sampling distributions of factor loadings. Psychom.33
- Rao,C.R.(1955): Estimation and test of significance in factor analysis. Psychom.20
- Schiller,U.& Wittmann,W.W.(1971): Monte-Carlo-Untersuchungen zur Image-Analyse. Zulassungsarbeit zur Diplomspsychologen-Hauptprüfung, Freiburg, unveröffentlicht
- Sixtl,F.(1967): Meßmethoden der Psychologie. Weinheim
- Tucker,L.R.,Koopman,R.F.,Linn,R.L.(1969): Evaluation of factor analytic research procedures by means of simulated correlation matrices.Psychom.34
- Oberla,K.(1968): Faktorenanalyse. Berlin
- Walter,E.(1970): Statistische Methoden I und II. Berlin/Heidelb./N.Y. (Hrsg)
- Winer,B.J.(1962): Statistical principles in experimental design. New York
- Winer,B.J.(1968): The error. Psychom.33
- Wittmann,W.W.(1976): Kaiser's Maß für Stichprobenadäquatheit (MSA): Bedeutung und Anwendungsmöglichkeiten für Faktorenanalyse und Testkonstruktion. Z.Exp.angew.Psychol. 23
- Zurmühl,R.(1964): Matrizen und ihre technischen Anwendungen. Berlin

Tabellarischer Lebenslauf

5.10.1940	Geboren in Magdeburg
1959	Abitur am Kepler-Gymnasium in Ulm/Do
1959 - 1962	Studium der Mathematik und Physik an den Universitäten Tübingen und München
1962 - 1964	Studium an der Pädagogischen Hochschule in Schwäbisch Gmünd mit Abschluß durch die 1. Dienstprüfung
1964 - 1967	Lehrer in Lörrach und Kandern
1967 - 1972	Studium der Psychologie an der Universität Freiburg mit Abschluß durch die Diplomprüfung
1972 - 1978	Assistent an der Pädagogischen Hochschule Karlsruhe
seit 1978	Wissenschaftlicher Angestellter an der Pädagogischen Hochschule Freiburg

